European Association for
Computer-Assisted Language Learning

EUROCALL 2023. CALL for all Languages

Short Papers

15-18 August 2023. University of Iceland, Reykjavik

edUPV
Universitat Politècnica de València

EUROCALL 2023

CALL for all Languages

Edited by

Branislav Bédi, Yazdan Choubsaz, Kolbrún Friðriksdóttir,
Ana Gimeno-Sanz, Súsanna Björg Vilhjálmsdóttir and Sofiya Zahova

edUPV
Universitat Politècnica de València

Congress UPV
EuroCALL 2023. CALL for all Languages

The contents of this publication have been approved by the Congress Scientific Committee and in accordance to the procedure set out in http://ocs.editorial.upv.es/index.php/EuroCALL/EuroCALL2023/about/editorialPolicies

First edition, 2023 (electronic version)

ISBN: 978-84-1396-131-6 (electronic version)

DOI: http://dx.doi.org/10.4995/EuroCALL2023.2023.17389

Table of contents

PEER-REVIEW COMMITTEE V

PREFACE
Branislav Bédi and Yazdan Choubsaz VII

APPLICATION OF THEORIES AND FRAMEWORKS SUPPORTING CALL
Generative AI tools in CALL: what are the options for teachers and language practitioners?
Branislav Bédi, ChatGPT C-LARA-Instance, Manny Rayner and Annika Simonsen 1

AUTOMATED CALL SOLUTIONS
An investigation of the efficacy of a ChatGPT-powered chatbot
Michael Wilkins, Mark Donnellan and Kane Linton 7

A process-based perspective on the effects of machine translation on L2 writing
Margot Fonteyne, Joke Daems, Maribel Montero Perez and Lieve Macken 13

A Rasch analysis validation of a survey on the use and beliefs of machine translation
Suwako Uehara 19

Assessing Google Translate ASR for feedback on L2 pronunciation errors in unpredictable sentence contexts
Paul John, Carol Johnson and Walcir Cardoso 25

ChatGPT and language learning: University educators' initial response
Louise Ohashi and Antonie Alm 31

Exploring users' perceptions of ASR for writing narrative texts
Danial Mehdipour-Kolour and Walcir Cardoso 37

Machine translation in language education: Perspectives from advanced language learners
Antonie Alm and Yuki Watanabe 43

Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice
Kevin Papin and Walcir Cardoso 48

COLLABORATIVE LEARNING AND CALL
An interdisciplinary telecollaboration in ESP classroom: a study involving Business and Law students
W A P Udeshinee and Carmenne Kalyaniwala 54

Cultural identity construction and English as a lingua franca in online virtual exchange
Alice Gruber and Sofiya Karnovska 61

Enhancing university EFL students' informal reasoning on the social scientific issues related to sustainable development goals by adopting a collaborative argumentation CALL environment
Li-Jen Wang 68

DISTANT LEARNING AND CALL

Collegial evaluation of online English for Specific Purposes (ESP) courses
Christopher Allen and Maria del Carmen Boloña 74

How do virtual reality lessons affect EFL learners' perceptions of speaking English?
Yoshiho Satake, Shinji Yamamoto and Hiroyuki Obari 80

ENDANGERED LANGUAGES AND CALL

App-assisted language revitalization: Insights from applied cognitive linguistics
Christina Ringel and Theresa Pohle 86

Harnessing the power of images in CALL: AI image generation for context specific visual aids in less commonly taught languages
Liang Xu, Elaine Uí Dhonnchadha and Monica Ward 92

The teacher-in-the-loop: collaboration with the teaching community in the development of an iCALL platform for Irish
Madeleine Comtois and Neasa Ní Chiaráin 98

HIGH-TECH AND LOW-TECH ENVIRONMENTS IN CALL

Listening tasks in virtual reality: A pilot study
Regina Kaplan-Rakowski and Yongluan Ye 104

Machine translation as a potential tool for comprehension in multilingual classes
Marni Manegre, Mar Gutiérrez-Colón and Andreu Martí Aguilar 111

INDIGENOUS LANGUAGES AND CALL

Evaluating the effectiveness of Microsoft Transcribe for automating the assessment of pronunciation in language proficiency tests
Carey Nelson and Walcir Cardoso 117

LANGUAGE CORPORA AND DATA-DRIVEN LEARNING

A pilot study of data-driven learning approach in teaching Chinese vocabulary
Yu-Ting Tseng and Li-Ping Chang 123

Data-driven learning beyond English: Insights and implications from three monographs
Luciana Forti, Nina Vyatkina and Eva Schaeffer-Lacroix 128

Japanese readability assessment using machine learning
Tyler Ivie and Robert Reynolds 133

Learner corpora and the design of data-driven learning activities
Luciana Forti 139

Making ICALL's intelligence accessible and understandable: evaluation, validation, and future directions of ICALL ecosystems
Jasper Degraeuwe and Patrick Goethals 145

LEARNER ASSESSMENT IN CALL

Developing LEMI: A new corpus based literacy support tool for schoolchildren
Roxana Rogobete, Alexandru Oraviţan, Mădălina Chitez and Karla Csürös 153

LESS COMMONLY SPOKEN LANGUAGES AND CALL

A machine-learning approach to Czech readability
Peter Williams and Robert Reynolds 159

Co-creating CALL content - does it work? Goldilocks compromise or Cruella chaos?
Monica Ward, Elaine Uí Dhonnchadha, Jennifer McGarry and Liang Xu 165

Didactization of L2 French variation in a complex context
Alena Barysevich 171

Resourceful approaches in CALL for less-commonly taught languages (LCTLs): Case studies on Icelandic, Irish, and Nawat
Jane O'Toole, Branislav Bédi and Monica Ward 177

Using Pimsleur for the self-regulated learning of spoken phrases in Brazilian Portuguese: a case study
Francis Will and Walcir Cardoso 182

Widening the lens on technology-mediated Irish language learning policy and practice to enable Irish language, technology, and education policy alignment, integration, and enactment at school level
Jane O'Toole 188

PROMOTING THE LEARNING OF LANGUAGES WITH CALL TECHNOLOGIES

Appropriating Zoom to host and conduct writing center tutorials
Chris Harwood 195

Beyond the walls of classrooms: Exploring the pedagogical effectiveness of Text-To-Speech-based Shadowing (TTS-S) on the development of Mandarin tones
Sue-Anne Richer and Walcir Cardoso 200

Cognitive grammar hands-on: a design-based approach to the didactic integration of interactive grammar animations
Daniel Pust 205

Effects of online task-based peer interaction on learners' speech development and attitudes toward English as a lingua franca
Atsushi Iina and Brian Wistner 211

Effects of text-to-speech synthesized speech on learners' presentation anxiety and self-efficacy: A comparison of two models
Takatoyo Umemoto, Shinnosuke Takamichi, Yuta Matsunaga, Yusuke Yoshikawa, Kikuko Yui, Kishio Sakamoto, Shigeo Fujiwara, and Yasushige Ishikawa 218

Enhancing pronunciation instruction for non-francophone immigrants in Quebec: a technology-assisted proposal
Geneviève Bibeau and Walcir Cardoso 223

Enhancing short academic presentations through extended independent practice using VoiceThread
Heejin Chang and Scott Windeatt 229

Exercise parameters influencing exercise difficulty
Tanja Heck and Detmar Meurers 236

FOCUS - For a functional, digital, and critical literacy
Beatrice Johansson and Qarin Franker 242

Investigation of PoodLL ReadAloud in Moodle to reduce the foreign language learning anxiety in English and Japanese language classes
Maki Terauchi Ho 249

Machine translation use in the
English as a Foreign Language (EFL) classroom
Anthony Young 254

Support inclusive teaching and learning with CALL technologies in the beginning-level Chinese language classroom
Lisha Xu 266

The use of machine learning in developing learner-adaptive tools for second language acquisition
Maryam Sadat Mirzaei and Kourosh Meshgi 272

Using generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic
Annika Simonsen and Branislav Bédi 278

Virtual reality: “Awesome”, “OK”, or “Not so good” for language learning?
Tricia Thrasher, Regina Kaplan-Rakowski, Dorothy Chun and Randall Sadler 293

Writing with automatic speech recognition: Examining user’s behaviours and text quality (lexical diversity)
Walcir Cardoso and Danial Mehdipour-Kolour 300

REGIONAL PERSPECTIVES ON DIGITAL LITERACIES

Assessing 21st century digital literacies in Japanese higher education
Sandra Healy and Olivia Kennedy 305

Attitudes of part-time Japanese university teachers on technology use after emergency remote teaching
Frances Shiobara, Kym Jolley and Mark Donnellan 311

Persian FLAIR: grammatically intelligent web search for language learning
Evan Bartholomeusz and Robert Reynolds 316

SYNCHRONOUS AND ASYNCHRONOUS LEARNING

Are digital learning paths sufficient for learning/practising English grammar? Views of secondary school pupils
Eirini Busack 322

Effective conversational practice in an Icelandic LMOOC
Kolbrún Friðriksdóttir 328

TEACHER TRAINING IN CALL

Language teacher training in CALL: fostering engagement in an online learning environment
Giovanna Carloni 333

Teacher training perspectives for virtual exchange: Initial actions from the E-LIVE European project
Ciara R. Wigham and Kristi Jauregi-Ondarra 339

PEER-REVIEW COMMITTEE

Allen, Christopher Michael; Linnaeus University, Sweden
Barysevich, Alena; University of Guelph, Canada
Boloña, Maria del Carmen; Universidad Casa Grande, Ecuador
Busack, Eirini; Karlsruhe University of Education, Germany
Cai, Kui; University of California, USA
Cardoso, Walcir; Concordia University, Canada
Carloni, Giovanna; University of Urbino, Italy
Chang, Li-ping; National Taiwan University, Taiwan
Chun, Dorothy; University of California, USA
Degraeuwe, Jasper; Ghent University, Belgium
Donnellan, Mark; Kindai University, Japan
Fonteyne, Margot; Ghent University, Belgium
Forti, Luciana; University for Foreigners of Perugia, Italy
Gruber, Alice; Augsburg Technical University of Applied Sciences, Germany
Harwood, Chris; Sophia University, Japan
Healy, Sandra; Kyoto Institute of Technology, Japan
Heck, Tanja; University of Tübingen, Germany
Ishikawa, Yasushige; Kyoto University of Foreign Studies, Japan
John, Paul; Université du Québec à Trois-Rivières, Canada
Johnson, Carol; Concordia University, Canada
Kaplan-Rakowski, Regina; University of North Texas, USA
Macken, Lieve; Ghent University, Belgium
Magana-Redondo, Juan Jose; UNED, Spain
Manegre, Marni Lynne; Universitat Rovira i Virgili, Spain
Markkanen, Marjo Helena; Kuopio Community College, Finland
Nelson, Carey; Concordia University, Canada
O'Toole, Jane Mary; Trinity College, Ireland
Ohashi, Louise; Gakushuin University, Japan
Oravițan, Alexandru; West University of Timisoara, Romania
Pust, Daniel; Otto von Guericke University Magdeburg, Germany
Reynolds, Robert; Brigham Young University, USA
Rogobete, Roxana; West University of Timisoara, Romania
Satake, Yoshiho; Aoyama Gakuin University, Japan
Shiobara, Frances Jane; Kobe Shoin Women's University, Japan
Terauchi Ho, Maki; Obihiro University of Agriculture and Veterinary Medicine, Japan
Thrasher, Tricia Kelly; Immerse, USA
Udeshini, Piyumi; Stockholm University, Sweden/NSBM Green University, Sri Lanka
Uehara, Suwako; The University of Electro-Communications, Japan

Preface

Branislav Bédi[a] and Yazdan Choubsaz[b]
[a]Division of Icelandic, the Árni Magnússon Institute for Icelandic Studies, , branislav.bedi@arnastofnun.is and [b]Department of English for Sciences, Center for Preparatory Studies, Sultan Qaboos University, , y.choubsaz@squ.edu.om

How to cite: Bédi, B.; Choubsaz , Y. (2023). Preface. In B. Bédi, Y. Choubsaz, K. Friðriksdóttir, A. Gimeno-Sanz, S. Björg Vilhjálmsdóttir & S. Zahova (Eds.), *CALL for all Languages - EUROCALL 2023 Short Papers.*

The 2023 EUROCALL conference was held for the second year in a row in Reykjavik on the 15th-18th of August 2023 but this time, after three years of online conferencing, as an in-person event hosted by the VIC – Vigdís International Centre, the Vigdís Finnbogadóttir Institute for Foreign Languages at the University of Iceland, and the Árni Magnússon Institute for Icelandic Studies. The conference programme included one full workshop day followed by a three-day agenda with multiple parallel sessions.

The focus of this year's EUROCALL conference was on inclusiveness of all languages, which was also reflected in the main conference theme, *CALL for all languages*.

The conference theme was central to all co-hosting institutes, particularly, however, the Vigdís International Centre – VIC, whose research and activities focus on the promotion, preservation, awareness, and support in education and research of foreign languages and cultures. The Centre works under the auspices of UNESCO and actively participates in the United Nations International Decade of Indigenous Languages (2022 to 2023). The role of the Árni Magnússon Institute for Icelandic Studies is to conduct research in the field of Icelandic studies and related scholarly disciplines, in particular Icelandic language and literature, and among other activities also to disseminate knowledge in these research fields. As such, the conference theme was dedicated to indigenous, endangered, and less commonly spoken languages in the Nordic Region, Europe, and the rest of the world, and welcomed submissions related both to the main conference theme as well as to other themes related to language technologies in CALL.

The 2023 EUROCALL conference engaged about 200 speakers from 30 different countries. Based on experience from the previous three years of online conferences, we also made use this year of the Whova® platform for communication and sharing photographs or other information, uploading additional documents and video presentations to sustain the EUROCALL community, allowing for exploration of content prior to and after the conference, sending online polls, publishing the official conference programme, and for staying connected during different strands:

- four workshop sessions;
- four keynote sessions;
- 47 group sessions in which three presenters gave a twenty-minute talk with a total of 141 presentations;
- three European project meetings with individual sessions in which each presenter gave a thirty-minute presentation;
- five symposia in which three to five speakers gave short presentations in each symposium;
- a poster session with seventeen poster presentations;
- nine special interest group meetings; and
- an editors' workshop.

The four keynote presentations reflecting the conference theme from various angles were given by the following invited speakers.

Júlía Guðný Hreinsdóttir, head of teaching Icelandic Sign Language – ÍTM at The Communication Centre for the Deaf and Hard of Hearing, delivered a talk entitled 'The history of sign language and the development of sign language teaching in Iceland'. In her talk, she demonstrated that Sign languages, like all other languages, differ from country to country. Icelandic sign language (ÍTM) is an example of a spontaneous language that developed without exposure to other languages, in a school for deaf students. Similar processes occur everywhere in the world: when deaf students come together in deaf schools they communicate in sign language. "Döff" is a language-minority group in Iceland, united by a special and shared history and culture and, most importantly, a communal natural language – ÍTM. The hearing status is of no importance. In June 2011 it was confirmed by the Icelandic Parliament, that ÍTM has equal status to Icelandic as a form of communication in Iceland. In the same act it was also declared that Icelandic is the national language of Icelanders and the official language of Iceland. In a similar way ÍTM is the 'national' language of the approximately 300 Döff individuals in Iceland.

Neasa Ní Chiaráin, ussher assistant professor at Trinity College Dublin, delivered a talk entitled 'Endangered Languages from an Irish perspective: *Tír gan Teanga, Tír gan Anam'*. In her presentation, she discussed the situation of Irish (Gaeilge) and the efforts currently underway, particularly those involving CALL, that aim to contribute to the revitalisation and maintenance of the language. She highlighted the importance of their indigenous language to their identity, which is underlined in the well-known saying by the Irish patriot Pádraig Pearse, "Tír gan Teanga, Tír gan Anam" – a country without its language is a country without its soul. However, she mentioned that the revitalisation of an endangered language is a daunting task and outlined some of the challenges where the development of digital resources and CALL can have a positive impact. Overall, her talk illustrated some of the general principles that inform her own work in iCALL, the research and development in ABAIR, which are enshrined in the Digital Plan for Irish Speech and Language Technology. These principles are to embrace the Irish indigenous language communities; to share ideas and resources with other endangered language communities; and to proceed with applications that harness whatever limited resources one has access to at a given point in time.

Róbert Bjarnason, CEO of the Citizens Foundation[1], delivered a talk entitled 'Leveraging AI for Language Education: An Overview of ChatGPT in Language Learning'. In his presentation, he demonstrated how language education is rapidly evolving, and how the integration of Artificial Intelligence (AI), particularly language generation models such as OpenAI's ChatGPT, can revolutionize the way we learn and teach languages. In his talk, he illustrated how ChatGPT can be be leveraged to enhance language learning outcomes and provide innovative solutions for language learners of all levels. From conversational agents that simulate real-life language interactions to personalized language learning materials, he demonstrated that ChatGPT has indeed the potential to transform the language learning experience. With advanced algorithms and Natural Language Processing (NLP) techniques, ChatGPT can provide customized and interactive language learning experiences, real-time feedback on pronunciation and grammar, and engaging dialogues.

Trond Trosterud, professor of Language technology at the Department of Language and Culture at UiT - The Arctic University of Norway, delivered a talk entitled 'CALL for all languages? Why languages differ and what consequences that has for CALL'. In his talk, he argued that organised foreign language teaching is typically done for the following two kinds of languages. The first is the handful of foreign languages found in school curricula (in various parts of Europe, traditionally German, English, Russian, and French). The second group consists of a large share of the approximately 100 official state languages. Both cultural contacts and work migration has, for these languages, given rise to production of learning material and more recently also CALL applications. Between 98-99% of the world's languages fall outside the two groups mentioned. They are either widely spoken but otherwise marginalised languages of former colonies, or they are languages of indigenous

[1] https://www.citizens.is

peoples. He pointed out that in order to provide both basic tools for a functioning literacy, philological insight, and language learning programs, we need to be able to model the vocabulary of the language, be able to analyse the grammatical properties of any wordform, and determine whether it belongs to the (standard or non-standard) language or not.

We would like to thank all the participants, keynote speakers, session presenters, poster presenters, members of the EUROCALL special interest groups, and workshop leaders for their invaluable contributions to the conference in general, and to the advancement of CALL, in particular.

Our extended gratitude goes to the EUROCALL Executive Eommittee for their constant support, the Conference Programme Committee, and last but not least to the local organising team, including our colleagues and student helpers.

This volume includes 55 selected short papers delivered at the EUROCALL 2023 Conference and it offers a combination of research studies and theoretical perspectives organized into 14 thematic strands. We would like to thank both the authors and the reviewers for their time and effort in ensuring that high scientific standards have been met in delivering this volume. Finally, we would like to warmly thank Ana Gimeno-Sanz and Sylvie Thouësny for their support in ensuring a smooth transition of the publication process from Research-publishing.net to the publication of the EUROCALL short conference papers with the Universitat Politècnica de Valencia's Open Conference System.

EUROCALL 2023: CALL for all languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16999

Generative AI tools in CALL: what are the options for teachers and language practitioners?

Branislav Bédi[a], ChatGPT C-LARA-Instance[b], Manny Rayner[b] and Annika Simonsen[c]
[a]Division of Icelandic, Árni Magnússon Institute for Icelandic Studies, branislav.bedi@arnastofnun.is; [b]UniSA STEM, University of South Australia, chatgptclarainstance@proton.me, Manny.Rayner@unisa.edu.au and [c]Department of Icelandic and Comparative Cultural Studies, University of Iceland, ans72@hi.is

How to cite: Bédi, B.; ChatGPT; Rayner, M.; Simonsen, A. (2023). Generative AI tools in CALL: what are the options for teachers and language practitioners? In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16999

Abstract

We present an exploration of generative artificial intelligence's potential in Computer-Assisted Language Learning (CALL), describing recent and expected near-future developments, tools, practical applications, and ethical considerations in the field. We trace different stages of development in CALL and argue that the advent of generative AI has inaugurated a new stage of Intelligent CALL. We consider the possible capabilities and limitations of AI as found in recent studies, compare with our personal, hands-on experience of using generative AI in CALL, and suggest ways to successfully utilise this technology. In particular, we briefly present the C-LARA project, which is a current focus of our activities.

Keywords: *artificial intelligence, new periodisation of CALL, teacher training, ethical considerations.*

1. Introduction

In this paper, our focus will be to examine the ongoing shift from traditional AI towards the new generative AI paradigm, as it applies to CALL. We offer a brief overview of how recent generative AI technology can be utilised in CALL and outline its capabilities and limitations. Traditionally, AI has been associated with Data-Driven Learning (DDL), corpus research, machine-learning, and areas of language technology including Text-To-Speech (TTS) and Automatic Speech Recognition (ASR): generative AI is rapidly encroaching on all of these, with corresponding consequences for CALL. The rest of the article is structured as follows: a brief description of AI and stages of evolution in CALL; exploring possible capabilities and limitations of generative AI in CALL; and finally, discussion and future directions.

2. A brief evolution of CALL

The evolution of CALL has been marked by distinct stages, with Warschauer (2000) identifying three historical phases: structural (1970s–1980s), communicative (1980s–1990s), and integrative (21st Century). Bax (2003) proposed alternative stages: restricted (1960s-1980s), open (1980s-2000s), and integrated CALL (2000s-present), with the last stage bringing CALL to a so called normalization, where technology becomes ubiquitous (Bax, 2003). This final stage would represent a major transition/paradigm shift (Lv, 2023), rendering technology

invisible in everyday teaching; this stands in sharp contrast to earlier approaches, where technology is viewed as supporting language education (Nyns, 1989; Shadiev & Wang, 2022), even in advanced applications like personalized learning through data analysis and instant feedback (Kohnke et al., 2023). Traditional AI relies on pre-existing data for pattern recognition, predictions, and classification. In contrast, generative AI also uses its training data to create original content, opening new perspectives for teaching and learning. Developing new digital skills, such as learning how to generate effective prompts (Kohnke et al., 2023), has now become part of Intelligent CALL. Technology has long supported language education and generative AI has brought about a new era, reshaping how we teach and learn in CALL.

3. Generative AI in CALL

This section explores how generative AI systems, e.g. ChatGPT-4[1], Bard[2], DALL-E-2[3], Midjourney[4], and Stable Diffusion[5], can be used for creating multimodal language learning resources. Please note that performance is strongly language dependent. Bard can currently only produce output in few languages. Even ChatGPT is currently much stronger in well-resourced, commercially important languages that have been prioritised by OpenAI. Performance in less-resourced languages is, however, improving quickly as more training data is deployed in training the models.

3.1. Capabilities of generative AI in CALL

With the above caveats, generative AI tools enable fast and relatively reliable ways of generating texts and images; as we write, audio capabilities are just about to be released for ChatGPT[6] and are promised for Google's Gemini[7]. Our observation is that the generative AI's current capability enables CALL teachers and practitioners to create useful tasks that help with creating new knowledge and content. Inspired by the most recent discussion and research (Bell, 2023; Crosthwaite & Baisa, 2023; Dilmegani, 2023; Ray, 2023), potential ways of using generative AI in CALL include the following:

1. Personalising lessons by learners' past performance, language skills, feedback, and learning abilities;
2. Organising, planning, and assessing lessons and course design to enhance individual learning experiences;
3. Creating and generating various learning materials, such as quizzes and flashcards, discussion questions, exercises, reading lists, study guides, summaries, video lecture scripts, language translations, grammatical correction or sentence analysis, and tagging text with parts of speech;
4. Generating new text and multimodal content;
5. Reviving outdated or poor-quality learning materials;
6. Offering tutoring through interactive chatbot discussions (conversational AI);
7. Developing digital skills for effective prompts and collaborative use of generative AI tools;
8. AIs collaborating with both technical and non-technical human experts to enhance CALL platforms; and
9. Integrating generative AI into linguistic and educational tools, supporting data-driven learning in real contexts, and Learning Management Systems (LMSs), transforming CALL to a more intelligent level, e.g. Duolingo Max[8], TinyStories with datasets of short stories created by ChatGPT-3.5 and ChatGPT-

[1] https://openai.com

[2] https://bard.google.com/

[3] https://openai.com/dall-e-2

[4] https://www.midjourney.com

[5] https://stablediffusionweb.com

[6] https://openai.com/blog/chatgpt-can-now-see-hear-and-speak

[7] https://lifearchitect.ai/gemini/

[8] https://blog.duolingo.com/duolingo-max/

4[9], C-LARA with integrated ChatGPT-4 functionality to generate texts, translations, and tagging (cf. section 3.3).

Harnessing these and related capabilities, the generative AI technology will soon offer a new generation of more advanced educational tools for language learning which will rapidly become part of CALL.

3.2. Limitations of generative AI in CALL

Although current research describes generative Large Language Models (LLMs) as revolutionary AI-driven technology, which has indeed a great potential in education (Kasneci et al., 2023; Kohnke et al., 2023), the same sources also voice various concerns:

1. As already noted, the fact that there is insufficient training data for all but the most highly prioritised languages means that the quality of generative AI tools is strongly language dependent.
2. Ethical aspects and copyright issues when using AI-generated output are important. The AI generates outputs guided by user prompts, but leveraging styles and genres taken from its training data (books, articles, paintings, audio recordings, videos), particularly if the material has been used without the original authors' consent, raises the possibility that some output can arguably be considered as plagiarised. Human editing of the output may alleviate the problem, but this both detracts from the usefulness of the generative AI technology and will not necessarily be accepted.
3. Other ethical and legal limitations include privacy and security issues (Bozkurt & Sharma, 2023; Kasneci et al., 2023; Kohnke et al., 2023; Short & Short, 2023).
4. Paid access to advanced versions of generative AI tools is another concern because monetisation limits access (Bozkurt et al., 2023). Language institutions and teachers have often limited resources that restrict them from using paid services for teaching and learning.
5. Shortage of training data may be manifested in cultural as well as linguistic problems. We ourselves found in a recent experiment with Icelandic (cf. Section 3.3 below) that cultural aspects from North America were often imported into Icelandic settings.
6. The widely remarked-on phenomenon of confabulations or 'artificial hallucinations' (Alkaissi H., McFarlane, 2023; Short & Short, 2023), endemic to generative AI, means that the factual aspects of generated content can in general not be considered reliable.
7. Lack of the necessary digital skills to generate effective prompts may affect the quality of the output.
8. Inequity and inequality are another issue to consider. Access to and using generative AI tools may be limited due to possible learners' disabilities and impairments (Kasneci et al., 2023).

3.3. Utilizing generative AI tools: our experience

Another study (Simonsen & Bédi, 2023) describes initial experiments with using generative AI tools to create multimodal texts directly. ChatGPT-4 was used to generate 15 anecdotal short stories for L2 Icelandic learners at various levels and ages. We added appropriate images using DALL-E 2, Midjourney, or Stable Diffusion. It was interesting to find that the same prompts used in the three tools gave substantially different results, and that the different AI systems required different amounts of detail in the prompts. In each case, the content creator could fine-tune the prompts by adding or changing wording to reach an image and style they personally liked.

A more substantial effort is currently under way in the shape of the ChatGPT-Based Learning And Reading Assistant (C-LARA; https://www.c-lara.org/; Bédi et al., 2023a, b). C-LARA takes inspiration from the earlier LARA platform (Akhlaghi et al. 2019) but represents a complete rewrite. Similarly to LARA, the basic goal is to develop a flexible platform that allows users to construct multimodal texts useful for L2 learners, but, as the name suggests, ChatGPT-4 stands in the center. The AI is used in two separate and complementary ways, both as

[9] Eldan and Li (2023). TinyStories: How Small Can Language Models Be and Still Speak Coherent English?. https://arxiv.org/pdf/2305.07759.pdf .

a software *engineer* and as a software *component*. In its software engineer role, ChatGPT has played a major part in developing the platform, collaborating closely with a human partner to write about 90% of the code and contribute the greater part of the software design.

In its software component role, ChatGPT can perform all the core language processing tasks. It can write short texts in response to user prompts, segment them into sentences and lexical units, annotate them with glosses and lemma/part-of-speech information, and add summaries. The prompts used to perform these operations are constructed from templates combined with few-shot sets of examples, facilitating language-specific tuning. The results are combined with TTS audio (the option of human-recorded audio will soon be added) and can be posted openly in a simple social network where other users are able to leave ratings and comments. The human annotator is able to edit the results of each stage, and there is functionality to compare different versions of the plain and annotated text, making it easy to estimate error rates for the different AI-based operations. For well-resourced languages prioritized by OpenAI, like English and Mandarin, ChatGPT-4 can write fluent, engaging texts in a wide variety of genres with an error rate of well under 1%; accuracy for glossing and tagging for these languages is typically in the mid-single digits, with the majority of the errors related to treatment of multi-word phrases. For less well-resourced languages, error rates can be much higher. A paper describing the issues in more detail, using a sample of five languages, is under review. We expect a public deployment of C-LARA to be available by the time this paper is published.

4. Summary and conclusions

With the utilisation of different AI technologies in CALL, a new era of Intelligent CALL has emerged in language education. We have tried to give a brief overview of the capabilities and limitations of generative AI. Generative AI tools offer many benefits for CALL, most importantly providing swift and reliable means for creation of new content. Potential applications range from personalised lesson planning and content creation to restoration and revival of outdated learning materials and tutoring. However, successful use of these tools typically requires effective input fine-tuning and prompt generation (Bozkurt, 2023; Kohnke et al., 2023; Short & Short, 2023; Kasneci et al, 2023). Use of generative AI in CALL has also raised concerns related to plagiarism, biases, privacy, security, equality, and equity, as well as careful consideration of ethical and copyright issues, and training data limitations mean that the tools so far only work well in the relatively small number of languages prioritised by the manufacturers. In conclusion, we briefly described our experiences with generative AI, in particular the collaborative open source C-LARA platform, where ChatGPT is used both as a software engineer and as a software component.

References

Alkaissi H. & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus 15*(2). PMC – PubMed. https://assets.cureus.com/uploads/editorial/pdf/138667/20230219-28928-6kcyip.pdf.

Akhlaghi, E., Bédi, B., Butterweck, M., Chua, C., Gerlach, J., Habibi, H., Ikeda, J., Rayner, M., Sestigiani, S., & Zuckermann, G. (2019). Overview of LARA: a learning and reading assistant. In *Proceedings of SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education*, 99–103. https://doi.org/10.21437/slate.2019-19.

Bax, S. (2003). CALL—past, present and future. *System 3*(1), 13–28. Elsevier. https://doi.org/10.1016/S0346-251X(02)00071-4.

Bell, E. (2023). Generative AI: How It Works, History, and Pros and Cons. Inverstopedia. https://www.investopedia.com/generative-ai-7497939.

Bédi, B., ChatGPT-4, Chiera, B., Chua, C., Ni Chiaráin, N., Rayner, M., Simonsen, A & Zviel-Girshin, R. (2023a). ChatGPT + LARA = C-LARA. Research Gate. https://www.researchgate.net/publication/373952306_ChatGPT_LARA_C-LARA.

Bédi, B., ChatGPT-4, Chiera, B., Chua, C., Ni Chiaráin, N., Cucchiarini, C., Rayner, M., Simonsen, A & Zviel-Girshin, R. (2023b). ChatGPT-Based Learning and Reading Platform: Initial Report. Technical report. ResearchGate.

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education 18*(1), 198-204. Asian Society for Open and Distance Education (ASODE), Japan. http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718/399.

Bozkurt, A., & Sharma, R. C. (2023). Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning. *Asian Journal of Distance Education 18*(21), i-viii. https://doi.org/10.5281/zenodo.7755273.

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., Toquero, C. M., Singh, L Tlili, A., Lee, K., Nichols, M., Ossiannilsson, E., Brown, M., Irvine, V., Raffaghelli, J. E., Santos-Hermosa, G Farrell, O., Adam, T., Thong, Y. L., Sani-Bozkurt, S., Sharma, R. C., Hrastinski, S., & Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53-130. https://doi.org/10.5281/zenodo.7636568.

Crosthwaite, P. & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics 3*, 1-4. Elsevier. https://doi.org/10.1016/j.acorp.2023.100066.

Dilmegani. C. (2023). op 6 Use Cases of Generative AI in Education. AIMultiple. Last update 22 April 2023. Accessed on 30 July 2023: https://research.aimultiple.com/generative-ai-in-education/.

Eldan, R., & Li, Y. (2023). TinyStories: How Small Can Language Models Be and Still Speak Coherent English?. https://arxiv.org/pdf/2305.07759.pdf.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G.. Michaeli, T., Nerdel, C.. Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J. Kuhn, J., Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences 103*. Elsevier. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868.

Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics 3*, 208-217. Elsevier https://doi.org/10.1016/j.cogr.2023.06.001.

Nyns, R. R. (1989). Is intelligent computer-assisted language learning possible?. *System 17*(1), 35-47. Elsevier. https://doi.org/10.1016/0346-251X(89)90058-4.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems 3*, 121-154. Elsevier. https://doi.org/10.1016/j.iotcps.2023.04.003.

Shadiev, R. & Wang, X. (2022). A Review of Research on Technology-Supported Language Learning and 21st Century Skills. *Frontiers in Psychology 13*. Frontiers. https://doi.org/10.3389/fpsyg.2022.897689.

Short, C. E. & Short, J. C. (2023). The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. *Journal of Business Venturing Insights 19*. Elsevier. https://doi.org/10.1016/j.jbvi.2023.e00388.

Simonsen, A. & Bédi, B. (2023). Using Generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic. In *B. Bédi, Y. Choubsaz, K. Friðriksdóttir, A. Gimeno-Sanz, S. Björg Vilhjálmsdóttir & S. Zahova (Eds.), CALL for all Languages - EUROCALL 2023 Short Papers*. https://doi.org/10.4995/EUROCALL2023.2023.16994

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16976

An investigation of the efficacy of a ChatGPT-powered chatbot

Michael Wilkins[a], Mark Donnellan[b] and Kane Linton[c]

[a]The Language Center, Kwansei Gakuin University, michaelwilkins@kwansei.ac.jp; [b]The Faculty of Informatics, Kindai University, donnellan@kindai.ac.jp and [c]School of Engineering, Kwansei Gakuin University, kane.linton.1985@gmail.com

How to cite: Wilkins, M.; Donnellan, M.; Linton, K. (2023) An investigation of the efficacy of a ChatGPT-powered chatbot. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16976

Abstract

One of the newest tools for language learning is AI-powered chatbots which allow students to engage in interactive conversations with a chatbot. English Central is a digital language learning platform that allows students to watch videos, study vocabulary, and produce spoken language through short interactive videos. In addition to these three core components, English Central has recently added a chatbot powered by ChatGPT. This paper shares the results of a pilot study that analyzed data from Japanese university students who engaged in interactions with this chatbot. The authors introduce quantitative data showing the quantity and quality of the interactions and discuss the strategies they implemented to encourage longer interactions. A student feedback survey was conducted and the results of this are also shared. The results suggested that although many students reported benefits to using a chatbot, the strategies implemented did not lead to significantly longer interactions, and that many students felt frustrated when the chatbot did not understand their utterances.

Keywords: *Chatbot, ChatGPT, English Central, AI*

1. Introduction

English Central (EC) is a digital language learning platform that allows students to do three core activities: watch videos; learn words; and speak lines from the videos. Language learners interact with these short videos getting invaluable listening and pronunciation practice and opportunities to learn vocabulary in context. In addition to these three core components, EC has recently integrated a chatbot called Mimi, powered by ChatGPT technology. For the purposes of this study, the researchers are specifically interested in students' interactions with Mimi. This paper reports on data from those interactions, shares student feedback on Mimi, and offers suggestions on how teachers can facilitate more meaningful interactions between chatbots and students.

Although the use of chatbots for language learning is considered by many to be a completely new field, there is already a well-established and rapidly growing body of supporting literature, including numerous review and meta-analysis articles (Fryer et al., 2020; Huang et al., 2022; Klímová & Ibna, 2023; Wollny et al., 2021; Zhang et al., 2023). Some researchers, such as Fryer & Carpenter (2006), have been working on using chatbots for language learning for almost 20 years. Key areas of inquiry into chatbot use have been: the affordances of chatbots (Wollny et al., 2021); instructional design and application (Fryer et al., 2020); learner objectives, experiences, and

challenges (Okonkwo & Ade-Ibijola, 2021); evaluation (Pérez et al., 2020); and technology design (Chen et al., 2023).

Chatbots are well-established as conversational agents to answer customer questions for online businesses and millions of people have interacted with digital assistants, such as Apple's Siri, showing a general acceptance of interacting with non-human entities. However, for language learning the results had not been as impressive (Fryer et al., 2020). With the recent advancements in AI, such as neural network language models, natural language processing, and the use of larger language model datasets, chatbots such as Open AI's Chat GPT 3.5 can interact with users in a more genuine and responsive way (Hong, 2023). However, to date, only a few studies have analysed the characteristics of the actual language produced by students while conversing with chatbots (Kim et al., 2021) and this study will attempt to address this gap.

ChatGPT and other AI tools are recognised to be massive disruptors of the language learning industry, but their effects are only just beginning to be felt (Hong, 2023). There are many articles assessing the pedagogical effects of chatbots, such as Klímová and Ibna (2023), but most are meta-analyses of the overall effect or evaluate a single use case rather than comparing different ways to integrate an existing chatbot into a classroom setting to achieve maximum student learning. There were no studies that looked specifically into the length and characteristics of actual student responses to chatbots in language learning. This study will explore optimal chatbot use in classroom settings.

A major concern for educators introducing new technology into the classroom is how students will react. If students do not perceive chatbots to be useful, entertaining, or both they will not use them except under compulsion from the teacher, which is not conducive to student learning. Huang (2022) asserts that chatbots can lower learners' affective filter and improve motivation as well as engagement in language learning. This study aims to expand on the findings that chatbots boost student motivation and comfort. The research questions were:

RQ 1: How long were typical student interactions with Mimi in terms of words per session?

RQ 2: What effect did the pedagogical interventions have on the interactions?

RQ 3: What are students' perceptions of the usefulness of chatbots?

2. Method

2.1. Context and participants

The participants were 26 students at a university in the Kansai region of Japan who used EC as a component of their bi-weekly four-skills English class during the spring 2023 semester. Each class in the 14-week semester was 100 minutes. The students used the three core components of EC, but also the EC chatbot Mimi was introduced.

2.2. Teaching interventions

The instructor took different pedagogical stances at different stages of the semester. For weeks 1 to 4 of the semester, students used EC without explicit pedagogical support from the teacher. They were given instructions, technical support, goals, and some encouragement to use Mimi, but no explicit 'teaching' was done. Marks were awarded for the core EC tasks of watching videos, vocabulary study, and pronunciation practice, but none for chatbot interactions. Students worked exclusively outside of class with just a short mention of the goals at the start and a reminder at the end of each class. In weeks 5-8, the teacher highlighted and praised the top-performing students and offered increased encouragement. A few marks were introduced for completing chats in addition to the regular marks for EC with work still being done exclusively outside of class time. In weeks 9-12, the process was like weeks 5-8 but the number of marks for completing chats was increased and extra marks were given for the quality of the chats. The teacher listened to one chat from each student and rated it from 1 to 5 points. The EC activity was still exclusively considered to be a homework task. In week 13 the teacher took a more direct role (intervention 1), and students were given significant time in class to watch EC videos and have conversations with

Mimi. In week 14 (intervention 2) the teacher gave a demonstration of a model chatbot interaction. It consisted of a five-turn chat with Mimi with each of the five 'student' turns being several sentences long and asked the students to follow that model. As in week 13, students were given significant time in class to do their EC video-watching and chat tasks.

2.3. Chatbot Data

The chatbot data for the whole semester was downloaded. This data included all utterances by both Mimi and the students, the number of turns in each interaction, and the average number of words per interaction. Mimi allows students to give both spoken and written responses.

2.4. Feedback Survey

Following the completion of intervention 2, a survey was given to gather student feedback on students' perceptions of EC, there were 13 6-point Likert scale items pertaining to chatbot usage, and qualitative data was collected through two open-ended questions to gather positive and negative student feedback on their interactions with Mimi. The student responses were coded thematically using NVivo.

3. Results

3.1. Chatbot Data

The data below shows a summary of the pertinent data from students' interactions with Mimi during three periods: pre-intervention 1; intervention 1 to intervention 2; and intervention 2 to the end of the semester.

Table 1. Summary of student interactions with the chatbot.

	Average no. of turns per session	Average no. of words per learner turn	Average no. of words per conversation	Percentage of spoken responses
Pre-intervention	4.67	10.85	50.66	57.9
Intervention 1 to intervention 2	4.94	10.56	52.16	63.9
Intervention 2 to end of semester	4.84	10.47	50.67	60.2

3.2. Survey results

The table below shows the responses to the 13 items on the survey that related to Mimi:

Table 2. Survey results.

	M	SD
Interacting with the English Central Chatbot helped me learn English.	3	1.5
English Central and its chatbot help English learning because they are available 24 hours a day.	2.3	1.4
English Central and its chatbot help English learning because they are available anywhere.	2	1.1
Speaking to the chatbot in English is less stressful than speaking to a human in English.	3.5	1.8

The chatbot interacted with me in a human-like way.	3.2	1.2
The chatbot seemed knowledgeable and did not say obviously false things.	2.6	1.2
The chatbot gave me good feedback on my English.	3.2	1.4
I could understand the chatbot easily.	2.5	1.3
I think speaking with the chatbot helped me speak English more fluently.	2.8	1.2
The teacher's request for students to use the chatbot motivated me to use the chatbot.	2.5	1.3
The only reason I used the chatbot was because I can get some points for this class.	2.7	1.4
The teacher using class time to explain the chatbot motivated me to use the chatbot more.	2.7	1.2
I will seek out opportunities to use chatbots to learn languages in the future.	3.2	1.2

Note. N=26 (1 completely agree – 6 completely disagree)

3.3. Open-ended responses

The students responded to two open-ended questions at the end of the survey about the positive and negative aspects of using the chatbot. The results of the thematic coding can be seen below in Table 3. Overall, there were 24 positive comments and 25 negative comments.

Table 3. Thematic coding of open-ended responses.

	Number of Comments
Positive	
Authentic Communication	1
Ease of Use	2
Generally Positive	12
Improved Skills (other than speaking)	2
Improved Speaking Skill	5
Reduced Anxiety	2
Negative	
Generally Negative	6
Inflexibility	1
Lack of Feedback	3
Miscomprehension	15

4. Discussion

Chatbots and AI language learning tools are relatively new, and teachers are examining how they can be best used for learning. This study set out to measure how much language students produce when interacting with chatbots, if teacher intervention increases this amount, and how students perceive this learning activity and tool.

The results show students took four to five turns of about ten words each for conversations of about 50 words in total per chatbot interaction. It is difficult to quantify this amount of student language production as sufficient or not. Since there are no studies to establish a previous baseline length or volume of student responses to chatbot prompts in language learning, an analysis of the results must be self-referential.

The results show that, regarding RQ1 and RQ2, the average number of turns increased and the average number of words per turn decreased slightly following the interventions. Also, following the interventions, the percentage of spoken responses increased. Given that the number of turns increased, the slight decrease in the word count is not surprising. The teacher interventions did not lead to significant changes in the number of words spoken by students per conversation with the chatbot, only slight improvements. There was no data in the survey comments to explain this, but it is possible that explicit teaching earlier in the semester would lead to more student output.

Regarding RQ3, student attitudes to chatbots were mixed. The positive aspect of Mimi always being available regardless of time and place was very popular and matches the consensus of previous research (Fryer & Carpenter, 2006; Kim et al., 2021) that one of chatbots' main affordances is their convenience regarding time and place. A second often-mentioned benefit was a lowering of the affective filter when interacting with chatbots (Kim et al., 2021; Klímová & Ibna, 2023). However, this is only partially reflected in the results. While some students found talking to Mimi less stressful with one student saying, "I am not nervous when I talked with him", others at the very least found the experience frustrating as is illustrated by the student comments regarding miscomprehension below. Chen et al. (2023) contend that chatbots may not understand student input or may not behave in a human-like way. The survey data supported this, with 15 students mentioning miscomprehension as being a major frustration when using Mimi. One student commented, "I hope their speech-to-text technology will be improved", with three others saying that Mimi "didn't catch my English correctly". While speech recognition technology has progressed and one student did comment that, "The chatbot can listen [sic] my English correctly", this student was in the minority with many of the responses suggesting that improved speech recognition is essential if Mimi and other chatbots are to become prominent tools in language learning.

5. Conclusions

This study showed that students were able to interact with Mimi and have a generally positive experience but with some complaints about miscomprehension and inflexibility in Mimi's speech recognition. This will most likely improve soon as the technology advances. Both positive and negative student comments are in line with the previous literature, so there seems to be a consensus forming regarding the issues surrounding chatbots in language learning. Surprisingly, teacher interventions did not greatly influence the quantity of student responses. In the next iteration of this research, the researchers hope to increase the number participants and discover what pedagogical interventions do increase student engagement with chatbots.

References

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, *10*(3), 8-14.

Fryer, L., Coniam D., & Carpenter, R. (2020) Bots for Language Learning Now: Current and Future Directions. *Language Learning & Technology*. 24(2) pp 8-22.

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1).

Huang, W., Hew, K. F., & Fryer, L. K. (2022). 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.12610

Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *영어학*, *21*, 712-734.

Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology*, *14*, 1131506. https://doi.org/10.3389/fpsyg.2023.1131506

Okonkwo, C. W., & Ade-Ibijola, A. (2020). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033.

Pérez, J. Q., Daradoumis, T., & Marquès Puig, J. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. https://doi.org/10.1002/cae.22326

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence*, *4*, 654924. https://doi.org/10.3389/frai.2021.654924

Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 1-21. DOI: https://doi.org/10.1007/s10639-023-11805-6

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16931

A process-based perspective on the effects of machine translation on L2 writing

Margot Fonteyne[a], Joke Daems[b], Maribel Montero Perez[c] and Lieve Macken[d]
[a]Ghent University, Ghent, Belgium, Margot.Fonteyne@UGent.be; [b]Ghent University, Joke.Daems@UGent.be; [c]Ghent University, Maribel.MonteroPerez@UGent.be and [d]Ghent University, Lieve.Macken@UGent.be

How to cite: Fonteyne, M.; Daems, J.; Montero Perez, M.; Macken, L. (2023). A process-based perspective on the effects of machine translation on L2 writing. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16931

Abstract

Most existing empirical work on the effects of Machine Translation (MT) use on second language (L2) writing has concentrated on its impact on writing products, with much less research addressing its effects on L2 learners' behaviours during writing. We therefore investigate whether the L2 writing process varies depending on whether learners are provided access to MT or to an online bilingual dictionary. Twenty-seven L1 Dutch learners of Swedish were assigned four writing tasks, with two tasks completed in each condition (MT or dictionary). While writing, participants' keystrokes were registered. Descriptive measures were used to summarize and compare participants' writing speed, tool engagement, and pausing behaviour across the two conditions. Results indicated that participants wrote more in less time, consulted the tool more frequently but for shorter periods of time, and paused longer between higher textual units when provided access to MT.

Keywords: *L2 writing process, machine translation, online dictionaries, keystroke logging.*

1. Introduction

The effects of MT use on L2 writing have gained increasing attention in recent years (Jolley & Maimone, 2022; Klimova et al., 2023; Lee, 2023). However, thus far, its effects have mostly been studied from a product perspective, overlooking the possible effects of using these tools on the processes learners engage in while writing a text in their L2. In this study, we adopt a process-based perspective to evaluate the impact of MT use on L2 writing. Our approach involved collecting, analysing, and comparing keystroke logs from 27 L2 writers, writing with access to MT or to an online bilingual dictionary. The L2 writing process was operationalised in terms of online measures of participants' writing speed, tool engagement, and pausing behaviour, which we extracted from the keystroke logs.

1.1. Background

L2 writers' linguistic knowledge is often limited and lacking automatisation (Kormos, 2012). Hence, the lower-order processes of vocabulary retrieval and grammatical encoding are commonly thought to impose greater cognitive demands on L2 writers, leaving them with less time and fewer resources to attend to higher-order processes, such as generating, organising, and evaluating the content of their texts. It has been suggested, however, that MT may increase L2 learners' lower-order processing speed. For example, Ahn and Chung (2020) found that respondents chose MT's time-saving capabilities most often as the main reason for using it during

writing. Similarly, in Clifford et al. (2013), respondents' third most common reason to use MT was that it "saves time" (p. 111). If MT indeed increases L2 learners' efficiency by accelerating their lexical retrieval and grammatical encoding processes, L2 writers should have more time and cognitive resources left to engage in higher-order processing, compared to when relying on more traditional tools such as online bilingual dictionaries.

Through triangulation of keystrokes with other data, studies have found associations between pauses and lower- and higher-order writing processes. Pauses between smaller textual units (within and between words) and larger textual units (between sentences and paragraphs) are typically linked to lower- and higher-order writing processes, respectively (Révész et al., 2019). In terms of duration, large thresholds such as 2000 ms are generally associated with higher-order processes, whereas smaller pause thresholds such as 200 ms are thought to also encompass lower-order processes (Van Waes & Leijten, 2015). Analysing L2 learners' pausing behaviour thus allows us to gather insights into how learners distribute their time and cognitive resources over higher- and lower-order processes during writing.

1.2 Research questions

Considering previous research on L2 learners' perceptions of MT and their writing behaviours, we formulated the following research questions:

1. Do L2 learners' writing speed and tool engagement vary depending on the tool they have access to while writing (i.e. MT or bilingual dictionary), indicating differences in the tools' efficiency?
2. Does L2 learners' pausing behaviour vary depending on the tool they have access to while writing, indicating differences in learners' distribution of time and resources over higher- and lower-order processes?

2. Method

2.1. Participants

A total of 27 learners of Swedish participated in this study. All participants exclusively identified Dutch as their first language. Out of the 27 participants, 21 learners were female, five were male, and one participant preferred not to disclose their gender. Participants were between 18 and 50 years old (M = 28.74, SD = 11.47, m = 22). Their L2 Swedish proficiency levels varied from A2 to B2 on the CEFR scale. At the time of data collection, 17 participants were enrolled in at least one (under)graduate course with Swedish as the main language of instruction at a Flemish university. The remaining ten participants were attending Swedish proficiency courses offered at various adult education centres in Flanders. All participants received monetary compensation (€12.50/hour) and were treated to a homemade Swedish pastry.

2.2. Materials

We developed four picture-based email writing tasks, all eliciting both descriptive and argumentative writing. Depending on the task, the participants were presented with images depicting three different events, holidays, travel destinations, or workshops. All images were connected to Sweden or its culture. For instance, in the 'event' task the images depicted an ice hockey cup, Sweden's national selection competition for Eurovision (*Melodifestivalen)*, and the Nobel Prize award ceremony in Stockholm. Participants were first asked to describe the three options presented to them using the images as references. Subsequently, they were instructed to provide arguments in favour of their chosen option and against the two other options. The emails had to be addressed to a (group of) friend(s) and written in Swedish. Participants were instructed to complete the four different writing tasks in Microsoft Word, with a maximum allotted time of 30 minutes for each task. Participants were told to aim for texts of 300 words, with descriptive and argumentative parts of equal lengths.

The study followed a within-subject design: participants were allowed to consult the MT tool *DeepL* for two of the tasks and were provided access to the online bilingual dictionary *Van Dale* for the other two tasks.

Participants were only allowed to translate between the language pairs Dutch-Swedish and Swedish-Dutch when using the tools. Usage of any other online or paper resources, including Word's built-in spelling and grammar checker, was prohibited.

2.3. Procedure

The procedure received ethical approval from the faculty. Each participant took part in two individual sessions, which lasted between 2.5 and 3 hours. Participants first provided informed consent and then engaged in practice tasks aimed at familiarizing them with the keyboard, software, and stimulated recall procedure. We assessed participants' L2 proficiency level and typing speed in Swedish using a proficiency test (Folkuniversitetet, n.d.) and a copy task (Van Waes et al., 2019), respectively. Additionally, participants completed a background questionnaire. Lastly, they were assigned two writing tasks. In the second session, participants first completed the remaining two writing tasks. Afterwards, participants took part in a stimulated recall session about their last writing task and were interviewed about their writing tool preferences and perceptions of the experiment.

To control for potential order, learning, and carry-over effects, we counterbalanced the order of the four writing tasks (events, holidays, destinations, or workshops) using a balanced Latin square design. In addition, we counterbalanced the order of the tools accessible to participants during the tasks (MT or dictionary). During each writing task, we registered participants' online writing behaviours using screen capture, eye tracking, and keystroke logging. For the keystroke logs we made use of Inputlog (Leijten & Van Waes, 2013).

2.4. Analysis

The 108 keystroke logs (four for each of the 27 participants) were analysed in terms of writing speed, tool engagement, and pausing behaviour. We obtained the writing speed measures by running a summary analysis on each log in Inputlog and extracting measures on participants' total process time (the time the participant needed to complete the task in minutes) and production rate (number of characters in the product, divided by the total process time). For the tool engagement statistics, we generated source analyses with Inputlog. Tool engagement was expressed with two measures: tool consultation frequency (number of times the participant opened the tool, divided by their total process time) and duration (mean duration of the times the participant had the tool open in seconds). Additionally, we used pause analyses generated by Inputlog to calculate pausing measures. Differences in pausing behaviour were studied in terms of total pause frequency (number of times the participant paused in Word, divided by the total time they spent in Word in minutes) and duration (median duration of their pauses in Word), and in terms of frequency and duration at different pause locations. To categorize the pauses by location, we employed Inputlog's classification scheme, distinguishing whether the pauses occurred within words, between words, or between sentences. We computed all frequency and duration measures twice, once with a 200 ms and once with a 2000 ms pause threshold. After extracting the writing speed, tool engagement, and pausing measures from each of the 108 keystroke logs, we grouped them by condition (MT or dictionary) and then calculated the mean values and standard deviations for each condition.

3. Results

For writing fluency, participants using the dictionary completed the tasks in an average of 28.63 minutes, at an average rate of 56.05 characters per minute (Table 1). The average participant in the MT condition exhibited a slightly shorter process time of 28.00 minutes and a slightly higher production rate of 58.59 characters per minute. Notably, the high standard deviations indicate considerable variability in writing speed across participants.

Regarding tool engagement statistics, participants consulted the dictionary approximately 0.99 times per minute, while for MT the frequency was slightly higher (1.06). The mean consultation length for the dictionary was 13.36 seconds, whereas this was notably shorter for MT (10.44s).

Table 1. Mean values and standard deviations (in parentheses) of the writing speed and tool engagement measures per condition.

Condition	Bilingual dictionary	MT tool
Total process time (in min)	28.63 (3.27)	28.00 (3.52)
Number of characters (per min)	56.05 (17.83)	58.59 (16.84)
Consultation frequency (per min)	0.99 (0.38)	1.06 (0.48)
Mean consultation duration (in s)	13.36 (3.96)	10.44 (3.70)

The results of the pause analyses for the two conditions are presented in Table 2. Mean values are given for all pauses and for pauses at three different pause locations, with two different pause thresholds. For pausing frequency, differences between conditions are small. In terms of pause duration, we do find that participants pause noticeably longer between sentences in the MT condition, especially when the larger pause threshold of 2000 ms is applied to the data. With MT, their median inter-sentence pause duration is 3.96 seconds on average, whereas they only pause between sentences for an average of 3.37 seconds with the dictionary.

Table 2. Mean values of the pausing measures per condition (BD: bilingual dictionary. MT: MT tool) by pause threshold (PT) in total and by location.

	PT	Total		By pause location					
				Within words		Between words		Between sentences	
Condition		BD	MT	BD	MT	BD	MT	BD	MT
Pause frequency (per min)	200	56.47	55.36	21.87	20.97	14.73	14.57	1.15	1.13
	2000	3.94	3.90	0.28	0.26	1.74	1.62	0.26	0.25
Median pause duration (in s)	200	0.41	0.41	0.29	0.28	0.68	0.65	0.98	1.13
	2000	3.30	3.31	2.63	2.66	3.21	3.27	3.37	3.96

4. Discussion and conclusions

In this paper, we presented a comparative analysis of the impact of two tools, an MT tool and an online bilingual dictionary, on the L2 writing process. We focused on differences in writing speed, tool engagement, and pausing. This section discusses the study's findings in relation to our research questions. However, it should be noted that these findings are only preliminary, as they are solely based on descriptive statistics.

Regarding our first research question, which addressed whether L2 learners' tool engagement and writing speed vary based on the tool they use, we observed that participants consulted the MT tool slightly more often. They also spent less time in it during a consultation. Furthermore, participants produced their texts at a slightly faster pace. These findings suggest that using MT may lead to a more efficient writing process. However, we also found high variability in participants' writing fluency. Considering previous research that has identified L2

proficiency as a strong predictor of fluency (Révész et al., 2022), the wide range of participants' L2 Swedish proficiency levels may explain this observation.

For our second research question, which investigated whether L2 learners' pausing behaviour differs depending on the available tool, we discovered that, when applying a pause threshold of 2000 ms, participants paused noticeably longer at higher textual units when using MT. Given that longer pauses between larger units are associated with higher-order processes, this finding suggests that participants focused more on higher-order concerns while writing with MT. However, this result needs to be validated through triangulation with other data sources.

Our objective is to address the limitations of this preliminary analysis in future research. First, we plan to construct linear mixed effects models for each of the measures presented. These models will allow us to make more robust generalizations about the impact of tool type on the L2 writing process and provide insights into potential moderating factors, including L2 proficiency level. Second, we will analyse the eye-tracking and stimulated recall data to explore differences in learners' reading behaviour during writing and the cognitive processes underlying learners' online writing behaviours. These analyses will yield additional insights into how learners allocate their time and cognitive resources to the various higher- and lower-order writing subprocesses. Finally, by annotating the eye-gaze, revision, and tool engagement behaviour in the screen recordings, we aim to uncover more fine-grained differences in L2 writers' patterns and the strategies they employ when using the two different tools.

Acknowledgements

This research has been carried out as part of a PhD fellowship on the WiLMa project, funded by Ghent University's Special Research Fund (Grant No. BOF.DOC.2021.0001.01). The first author also received a travel grant from the Research Foundation – Flanders (FWO) for her participation in the EUROCALL 2023 conference.

References

Ahn, S., & Chung, E. S. (2020). Students' perceptions of the use of online machine translation in L2 writing. *Multimedia-Assisted Language Learning*, *23*(2), 10–35. https://doi.org/10.15702/mall.2020.23.2.10

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *@tic. Revista d'innovació Educativa*, *10*, 108–121. https://doi.org/10.7203/attic.10.2228

Folkuniversitetet. (n.d.). *SVENSKA – Inplaceringstest B*. https://uiss.org/wp content/uploads/2017/08/INPLACERINGSTESTB.pdf

Jolley, J. R., & Maimone, L. (2022). Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal*, *14*(1), 26–44. https://doi.org/10.5070/L214151760

Klimova, B., Pikhart, M., Benites, A. D., Lehr, C., & Sanchez-Stockhammer, C. (2023). Neural machine translation in foreign language teaching and learning: A systematic review. *Education and Information Technologies*, *28*(1), 663–682. https://doi.org/10.1007/s10639-022-11194-2

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Lee, S.-M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *36*(1–2), 103–125. https://doi.org/10.1080/09588221.2021.1901745

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, *41*(3), 605–631. https://doi.org/10.1017/S027226311900024X

Révész, A., Michel, M., Lu, X., Kourtali, N., Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, *58*, 100927. https://doi.org/10.1016/j.jslw.2022.100927

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition*, *38*(Part A), 79–95. https://doi.org/10.1016/j.compcom.2015.09.012

Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with Inputlog. *Journal of Open Research Software*, *7*(1), Article 1. https://doi.org/10.5334/jors.234

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16954

A Rasch analysis validation of a survey on the use and beliefs of machine translation

Suwako Uehara[a]

[a]Graduate School of Informatics and Engineering, University of Electro-Communications, uehara.suwako@uec.ac.jp

How to cite: Uehara, S. (2023). A Rasch analysis validation of a survey on the use and beliefs of machine translation. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16954

Abstract

Many English as a Foreign Language (EFL) students use Machine Translation (MT) to a varying degree and studies on the students' use and beliefs of MT are primarily conducted through surveys and interviews. It is important for educators to understand students' use and beliefs related to MT, however, there are few studies in the Japanese setting, and none validated the surveys prior to implementation. The goal of this study was to validate a survey on MT to be used in the Japanese context. Considerations for validity are crucial when designing surveys for quantitative analysis. In this study, a 41-item Likert scale survey designed to understand students' opinions of MT use and beliefs in the Japanese tertiary education system was piloted with 93 first-year EFL learners in an academic writing class. Subsets of items targeting use, volume of text, and degree of acceptability were validated and optimized using the Rasch analysis. Results showed that further work is required to increase items for subsets and to reoptimize the instrument.

Keywords: *Rasch validation, machine translation, survey, use and belief.*

1. Introduction

In English as a Foreign Language (EFL) classrooms, Machine Translation (MT) tools such as Google Translate and DeepL are free and easily accessible to learners in Japan. Studies on the students' use and beliefs of MT are primarily conducted through surveys and interviews, and reports show MT is used by students for vocabulary learning, reading comprehension, and writing assignments (Jolley & Maimone, 2022). As these learners increasingly use MT tools for language learning and assessments, some instructors are concerned that MT could be harmful because learners may not be engaged in the writing process, and the use of MT could violate academic integrity (Vinall & Hellmich, 2021). It is therefore important for educators to understand students' use and beliefs related to MT in order to develop teaching strategies in the company of such readily available tools.

Okita and Kurokawa (2023) investigated machine translation use by graduate students in Japan through open-ended questionnaires and found graduate students use MT to read text in English, back translate their own writing, and as a dictionary. Those who did not use MT were wary of the quality of MT and others responded their written language ability in L2 was good enough not to rely on MT. There are still very few studies of student use and beliefs in the Japanese setting, and in addition, to my knowledge, none validate the surveys' items prior to implementation.

The goal of this study is to validate and refine a survey on MT that can be used in the Japanese context. The psychometric properties via the Machine Translation Use and Beliefs Survey (MTUB-S), a survey designed for this study, will be analyzed using Rasch analysis (Rasch, 1960) and implementing guidelines to refine scales

(Linacre, 1997) to fine-tune the rating scales. With this in mind, the research question is as follows:

RQ: To what extent does an instrument designed to measure Japanese university students' use and beliefs of machine translation in an EFL writing class meet the expectations of the Rasch model?

2. Method

2.1. Participants and educational context

Participants were Japanese L1-speaker science majors (N = 93; M = 85, F = 8; age 18–20) from a science and technology university located in Kanto, Japan. Participants were convenience sampled and recruited from three intact 1st-year EFL academic writing classes. English proficiency measured by TOEIC-IP (Institutional Program) scores taken in June 2023 ranged from 165 to 805, with an average score of 542. All participants were explained the purpose of the study and signed informed consent forms and data were collected in June 2023.

2.2. Instrument

The Machine Translation Use and Beliefs Survey (MTUB-S) is a list of 41 items comprised of two to six point Likert-scale items and multiple choice questions that are being developed to measure learners' use and beliefs of MT. Due to space restrictions, in this paper, I elaborate on Subsets Use (13 items) that are designed to measure the use of MT. Supplementary Data for volume of text, and beliefs on the degree of acceptability are available in Uehara (2023b). The aim of this study is to refine the MTUB-S items following the method described in the Analysis section. The items are available in bilingual format, and participants received the Japanese version. Following recommendations outlined by Messick (1989) on construct validity (content, substantive, structural, and generalizability), and Nemoto and Beglar (2014) for Likert scale item design, the 41 Likert scale items were generated by adapting items used in studies on MT, and by developing original items through a qualitative study by Uehara (2023a). The instrument was reviewed for feedback by four university instructors currently in a TESOL PhD program. The translations were back translated and reviewed for feedback by two bilingual tertiary level instructors who have translation experience. Polytomous Likert scales items ranging from 1 (e.g. *Never*) to 6 (e.g. *Always*) were used. The even numbered six point Likert scale format allows no neutral position and was chosen because it requires the respondents to provide an opinion that either agrees or disagrees to varying degrees (Krosnick & Fabrigar, 1997). The instrument was pre-tested with 20 students from the same institutions and the list was then reviewed and adapted with an expert in Rasch and Likert scale designs who has a PhD in TESOL. See Appendix A for the list of Subset Use. The survey instructions and full MTUB-S list can be found in the Supplementary Data S5 (See Uehara, 2023b).

The survey was disseminated online to the participants recruited from three intact academic writing classes, and they all responded to the survey during class time for about 10 minutes. The data were then subjected to Rasch analyses with WINSTEPS (Linacre, 2022). Students also responded to the open-ended prompt: "As a student attending English language classes, describe a situation that you think MT use is acceptable or unacceptable."

2.3. Analysis

The Rasch-Andrich rating scale model (Andrich, 1978) using guidelines from Linacre (1997, 2002) was used to follow best practices to optimise the Likert scale survey items. Rasch analysis is a statistical technique that can be used to analyze and refine surveys. The Rasch based approach places people (students) and items (each survey question) on a single hierarchical, equal interval logit scale. Not all items are of equal difficulty. An ideal set of items will include a range of items that are easy or more difficult to answer in order to examine the structure of a variable. Rasch will identify the separation between each item, and future considerations can be made to refine with new items to fill the gap in item difficulty. In addition, the advantage of this approach over reporting raw score averages and percentages is that the conversion to the logic scale is a calibration with fixed intervals, hence Rasch represents linearity across the respondents which justifies conclusions drawn from the data. Rasch can also identify unexpected responses from particular respondents (Tatum, 2000).

Linacre (1997) outlined guidelines for fine tuning rating scales. To ensure data accuracy first the dataset was thoroughly examined for any errors. The following values were then examined: reliability and separation; item and person fit [outfit mean square (MNSQ) less than 2.0); Wright map (to assess the impact of misfit items on targeting or expanding the item range); category probability curves (to detect irregular usage patterns); average category measure advance (to evaluate observed average measure advancement); and Andrich thresholds advance (to prevent disordered thresholds). The Andrich thresholds advance should be at least 1.4 logits and less than 5.0 logits to maintain an appropriate category width and avoid dead zones. Misfitting items were investigated and addressed by removing or collapsing them within the scale. Collapsing categories means the new dataset and responses are not truly representative of the respondents' responses. However, the choice of collapsing is practised when the model is subject to exploration (Wright & Linacre, 1992) and the researcher is responsible for making justified choices. The written responses in Japanese to the prompt regarding the acceptable or unacceptable use of machine translation were examined for any misfitting students. Following fit analysis, principal component analysis (PCA) of the Rasch residuals was conducted. According to Linacre (2022) if the eigenvalue of the unexplained variance in the first contrast is more than 2.0 the subset may not be unidimensional and item clusters comprising the contrast should be examined for substantive meaning. These analyses were repeated successively on WINSTEPS to fine-tune the survey items, generally removing items one by one. Results are shown in the next section.

3. Results

For all subsets, the data set was first run with all items in each subset separately. Each subset required at least four runs where misfitting items (infit over 1.5) were removed and categories collapsed. See Appendix B for the summary of results for Subset Use, VolTxt, and DegAcpt. Subset Plcy was not analysed for this study. See Suppmentary Data S1, S2, and S3 (MTUB-S Use, VolTxt, & DegAcpt) in Uehara (2023b) for the output of relevant tables and figures per subset.

The six step 13 item subset Use was optimized when reduced to 10 items by removing items 19, 20, and 10 and by collapsing the scale from 6 steps to 5 steps. In the first run, infit and outfit MNSQ underfit for item 19 (infit = 1.59; outfit = 1.79) and 20 (infit = 1.58; outfit = 1.59). Item 10 which was close to overfitting (first run infit = 0.58; second run infit = 0.54) was removed in the third run because "I use machine translation" was deemed to be a "summary item," (Sick, 2012) which is redundant and lacking independence from other items in the subset. Overall, separation and reliability improved somewhat for persons but decreased slightly for items. The Wright map indicated these items 19 and 20 did not help with targeting or extending the range. There was an irregularity in category observation (see Figure 1 Left). The probability curve showed reason to collapse categories 2 and 3 into a single category. The probability curve improved in the last run (See Figure 1 Right) observed average increased incrementally, however the Andrich threshold did not increase by 1.4 logits. The eigenvalue of the unexplained variance reduced from 3.64 to 2.77. Close inspection of the PCA standardized residual loadings in the last run revealed that items 14–18 seem to relate to editing, while items 11–13, 21 and 22 seem to relate to where and in what language MT is used (See Table 1). Finally, the average ability ascend improved, however one item (item 21) remained with an average ability that did not ascend with the category score. Student #18 scored very high and the highest for all categories in this subsection. The students' responses were checked but there was no strong evidence to remove this student. See Supplementary Data S6 in Uehara (2023b) for results and discussion of VolTxt and DegAcpt.

Table 1. Principle component analysis: standardized residual loadings for item of last run for subset use

CON-TRAST	LOAD ING	MEA-SURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY N0.	ITEM	LOAD-ING	MEA-SURE	INFIT MNSQ	OUTFIT MNSQ	ENTR Y N0.	ITEM
1	.75	-.27	.68	.67	A 11	11 Use_JtoE	-.63	-.66	.93	.95	a 16	16 Use_post-edit
1	.72	-.36	.64	.62	B 12	22 Use_at home	-.60	.16	1.20	1.17	b 17	17 Use_back trans
1	.56	.80	1.23	1.22	C 21	21 Use_in class	-.48	.23	1.39	1.41	c 15	15 Use_pre-edit
1	.54	-.47	.65	.70	D 12	12 Use_EtoJ	-.27	-.65	1.13	1.11	d 18	18 Use_satisfied
1	.10	.51	.80	.80	E 13	13 Use_ownJtoE	-.19	.70	1.26	1.21	e 14	14 Use_ownEtoJ

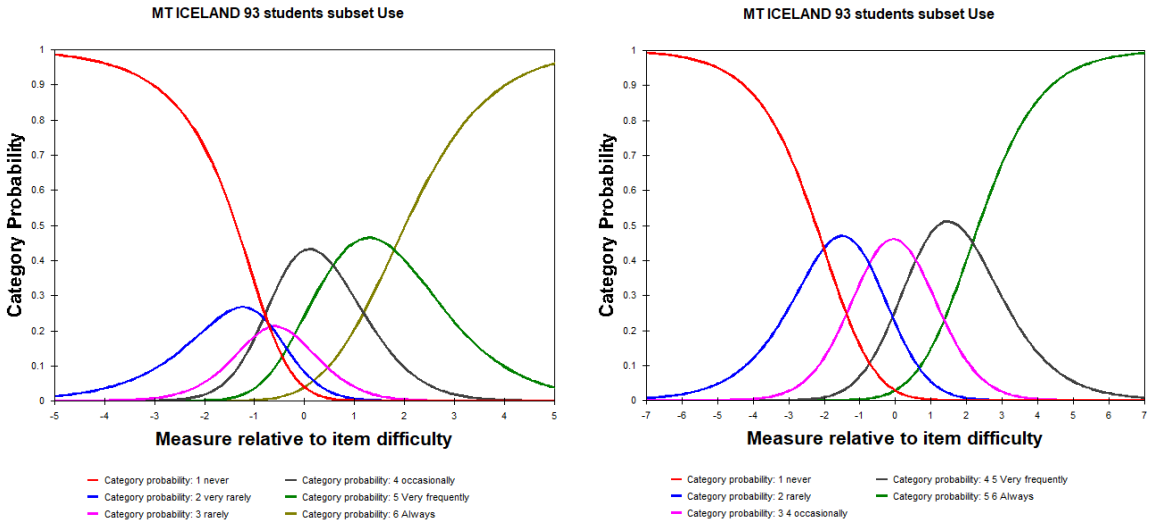


Figure 1. Category probabilities: Modes- Andrich thresholds at intersections (left = first run of subset use; right = last run of subset use)

4. Discussion and conclusions

In this study, Rasch analysis was conducted on subsets of MTUB-S as part of the validation of a new test instrument. For subset Use, the resulting 10 items on a collapsed 5-point step might consist of two dimensions: (1) how MT is used for editing, and (2) what mode (where and in what language direction) it is used. Item 19 (I use machine translation only in a way that benefits my language acquisition) and Item 20 (I use machine translation by considering how it can benefit my language acquisition) had poor fit, possibly because the phrase "language acquisition" was too specialized for the respondents, and were removed. The concept underlying Item 19 and Item 20 itself is an important one, so it will be meaningful to revive these two items through rewording in a future Rasch analysis. See Supplementary Data S6 in Uehara (2023b) for a discussion of VolTxt and DegAcpt.

Designing well-validated surveys (e.g. Messick, 1989) presents a greater challenge than one might anticipate. For this study more items need to be considered to make further improvements for separation and reliability, and dimensionality and the current list should be refined further. Future work will include adding more items through think aloud techniques and interviews with students. Such rigorously validated surveys can then be implemented across different studies to improve reliability and consistency, thereby enhancing the results of future studies.

5. Limitation

The data were convenience sampled from students at a science and engineering university which has a high percentage of male students. Therefore, the findings are not representative of all Japanese university students. Due to the limitation of space, misfiting students were not mentioned in this paper and future research will include data from a mixture of arts and science students across different universities and refine the survey items further based on the results of this study.

References

Jolley, J., & Maimone, L. (2022). Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal, 14*(1), 26–44. http://repositories.cdlib.org/uccllt/l2/vol14/iss1/art

Linacre, J. M. (1997). Guidelines for rating scales. In *Midwest Objective Measurement Seminar. Chicago: MESA Press, Research Note* (Vol. 2).

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85–106.

Linacre, J. M. (2022). *Winsteps® Rasch measurement computer program user's guide.* Version 5.2.3. Winsteps.com

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). Macmillan.

Nemoto, T., & Beglar, D. (2014). Developing Likert-scale questionnaires. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings.* Tokyo: JALT.

Niño A. (2020). Exploring the use of online machine translation for independent language learning. *Research in Learning Technology*, 28, 1–38. https://doi.org/10.25304/rlt.v28.2402

Okita, M. & Kurokawa, S. (2023). Machine translation and graduate students in Japan. *Komaba Language Association Journal 7*, 1–15.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Danmarks Paedigogiske Institut.

Sick, J. (2011). Rasch measurement in language education part 6: Rasch measurement and factor analysis. *SHINKEN: JALT Testing & Evaluation SIG Newsletter. March 2011*, *15*(1), 15–17. https://hosted.jalt.org/test/PDF/Sick6.pdf

Tatum, D. S. (2000). Rasch analysis: An introduction to objective measurement. *Laboratory Medicine*, *31*(5), 272–274.

Uehara, S. (2023a). Teacher perspectives of machine translation in the EFL writing classroom. In P. Ferguson, B. Lacy, & R. Derrah (Eds.), L*earning from Students, Educating Teachers—Research and Practice.* JALT2022, 270–279. https://doi.org/10.37546/JALTPCP2022-31

Uehara, S. (2023b, August 15–18). *A Rasch validation of a survey on machine translation use* [Paper Presentation]. EUROCALL 2023, Reykyavík, Iceland. https://doi.org/10.13140/RG.2.2.12717.87526

Vinall, K., & Hellmich, E. A. (2021). Down the rabbit hole: Machine translation, metaphor, and instructor identity and agency. *Second Language Research & Practice*, *2*(1), 99–118. http://hdl.handle.net/10125/69860

Wright, B., & Linacre, J. M. (1992). Combining (collapsing) and splitting categories. *Rasch Measurement Transactions*, *6*(3), 233–235. https://www.rasch.org/rmt/rmt63f.htm

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16954

Appendices

Appendix A

List of Items and Descriptors for Subset Use

Item No.	Item Descriptor for Subset Use
10	I use machine translation.
11	I use machine translation to translate text from Japanese to English.
12	I use machine translation to translate text from English to Japanese.
13	I use machine translation by writing my own text in Japanese and translating the text to English.
14	I use machine translation by writing my own text in English and translate the text to Japanese.
15	I pre-edit text that I intend to machine translate.
16	I post-edit text that I machine translated.
17	I back-translate the output from English to Japanese to check the initial translation using machine translation.
18	I check the content and use the text generated from machine translation only if I am satisfied with it.
19	I use machine translation only in away that benefits my language acquisition.
20	I use machine translation by considering how it can benefit my language acquisition.
21	I use machine translation in the classroom.
22	I use machine translation for assignments completed at home.

Note. Likert scale descriptors for subset Use are: 1 Never, 2 Very rarely, 3 Rarely, 4 Occasionally, 5 Very frequently, 6 Always; Item No. = Item numbers (Item 1 to Item 9 are a separate set of items measuring students' perceptions of their writing ability, and was not used in this study).

Appendix B

Summary of Subsets by Person and Item Separation, Reliability, Misfit, Point-Measure Correlation, Ascending Observed Average, Andrich Threshold Advance, Unexplained Variance, and Details of Data Set (No of Items (Items Removed), No of Steps, Code)

	Person		Item		Misfit*1							Data Set		
No	Sep	Rel	Sep	Rel	Infit	Outfit	PT Msr Corr *1	Item *1	Obsvd Avg Ascnd	And Thrs Adv	Unexp Vari	Item No (Items Rmvd)	Stp	Code
Subset Use														
1	1.99	.80	3.6	.93	1.59 1.58	1.79 1.59	.30 .37	19 20	Yes	No	3.6471	13	6	"123456"
2	2.25	.83	3.92	.94	NA	NA	NA	NA	Yes	No	3.1559	11 (19, 20)	6	"123456"
3	2.13	.82	3.82	.94	NA	NA	NA	NA	Yes	No	2.6745	10 (19, 20, 10)	6	"123456"
4	2.18	.83	3.66	.93	NA	NA	NA	NA	Yes	No	2.7774	10 (19, 20, 10)	5	122345
Subset VolTxt														
1	1.93	.79	1.81	.77	1.98	1.91	.51	23	Yes	No	2.6382	5	6	"123456"
2	2.25	.83	2.76	.88	1.65	1.5	.66	24	Yes	No	1.6715	4 (23)	6	"123456"
3	2.17	.82	2.67	.88	1.52	1.49	.7	24	Yes	Yes	1.7286	4 (23)	5	122345
4	2.43	.85	3.88	.94	NA	NA	NA	NA	Yes	Yes	1.6491	3 (23, 24)	5	122345
Subset DegAcpt														
1	2.82	.89	7.56	.98	1.60 1.50	1.56 1.47	.65 .69	37 36	Yes	No	2.8881	10	6	"123456"
2	2.69	.88	6.02	.97	1.84	1.99	.67	36	Yes	No	2.2023	9 (37)	6	"123456"
3	2.63	.87	3.59	.93	1.48	1.71	.68	32	Yes	No	2.1584	8 (37, 36)	6	"123456"
4	2.78	.89	3.67	.93	1.43	1.53	.73	32	Yes	No	2.2242	8 (37, 36)	5	122345

Note. Data Set = command file set to run the data set described; No = Command file run on WINSTEPS; Sep = Separation; Rel = Reliability; Infit = Infit MNSQ; Outfit = Outfit MNSQ; PTMsr Corr = Point-measure Correlation; Item = Item number; Obsvd Avg Ascnd = Whether observed average is in ascending order; And Thrs Adv = Whether the Andrich Threshold increments by more than 1.4 but less than 5.0 logits; Unexp Vari = Unexplained Variance; Item No (Items Rmvd) = Relevant item number for values in that run (Items that were removed in that run); Stp = No of steps; Code = Original in parenthesis, New codes out of parenthesis to collapse the scale; *1 = Report for those items which misfit > 1.5 or < 0.5.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16987

Assessing Google Translate ASR for feedback on L2 pronunciation errors in unpredictable sentence contexts

Paul John[a], Carol Johnson[b] and Walcir Cardoso[c]

[a]Université du Québec à Trois-Rivières, CogNAC Research Group, Trois-Rivières, Canada, paul.john@uqtr.ca; [b]Concordia University, Centre for the Study of Learning and Performance, Montreal, Canada, carol.johnson@concordia.ca and [c]Concordia University, Centre for the Study of Learning and Performance, Montreal, Canada, walcir.cardoso@concordia.ca

How to cite: John, P.; Johnson, C.; Cardoso, W. (2023). Assessing Google Translate ASR for feedback on L2 pronunciation errors in unpredictable sentence contexts. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16987

Abstract

Following previous research into predictable sentence contexts, this study assesses the pronunciation feedback provided by Google Translate's (GT) Automatic Speech Recognition (ASR) in unpredictable contexts. We examined the accuracy of GT transcriptions for target items recorded by male and female Quebec Francophones (QFs). The items occurred in neutral carrier sentences such that no contextual cues help ASR identify the targets. Th-initial vs t-initial (thank-tank) *and h-initial vs vowel-initial* (heat-eat) *items were used to investigate the potential for feedback on the QF errors of th-substitution, h-deletion, and h-epenthesis, comparing real-word* (thank→tank) *vs nonword output* (thief→tief). *As with predictable contexts in our previous research, we observed high transcription accuracy for real words only. Without contextual cues, accuracy rates were lower than in predictable contexts for correctly pronounced items but higher than for incorrect pronunciations constituting real words. Unpredictable contexts are thus inferior at confirming correct pronunciation (confirmative feedback) but superior at flagging real-word errors (corrective feedback). Contrary to the anticipated ASR gender bias, female recordings showed higher transcription accuracy than male recordings. Our findings both confirm the usefulness of GT's ASR for generating pronunciation feedback and highlight the importance of context (predictable vs unpredictable) and lexical status (real vs nonword).*

Keywords: *automatic speech recognition, Google Translate, L2 pronunciation, corrective vs confirmative feedback, predictable vs unpredictable contexts, gender bias.*

1. Introduction

The current study expands on an earlier investigation into using Google Translate's (GT) Automatic Speech Recognition (ASR) for *corrective* and *confirmative feedback* on second language (L2) pronunciation errors. Although corrective feedback can help learners improve pronunciation (Saito, 2021), questions remain regarding the accuracy of ASR feedback (Inceoglu et al., 2022). L2 pronunciation errors are generally variable, meaning learners alternate between targetlike and erroneous realizations of L2 sounds. Consequently, ASR transcriptions should not only flag incorrect pronunciation but also confirm correct pronunciation, hence our introduction of

the term *confirmative feedback* as a complement to *corrective feedback*. While our previous work investigated GT ASR transcription accuracy for items in predictable sentence contexts (John et al., in press), the current study assesses the technology in unpredictable contexts. The purpose is to establish the impact of presence vs absence of contextual cues on GT ASR's ability to provide feedback on correct and incorrect pronunciations.

We focus on three Quebec Francophone (QF) pronunciation errors in English: th-substitution (*thank→tank, thief→tief*), h-deletion (*heat→ _eat, help→ _elp*), and h-epenthesis (*old→hold, ice→hice*) (John & Frasnelli, 2022). Crucially, we investigate the role of lexical status (real vs nonword) and gender (male vs female voices) on transcription accuracy. While correct pronunciations always constitute real words, pronunciation errors can generate real or nonwords (*tank, _eat, hold* vs *tief, _elp, hice* above). Since nonwords are by definition absent from the GT lexicon, we anticipated high transcription accuracy for real words only. Nonetheless, unpredictable contexts should generate lower accuracy than correctly produced items in predictable contexts, where contextual (syntactic-semantic-collocational) cues conspire with phonetic cues to ensure correct identification of the target item (Ashwell & Elam, 2017). With respect to gender, ASR systems are often trained on datasets with more male speech samples, potentially leading to poorer performance on recognizing female voices (Garnerin et al., 2019). Gender bias in ASR would undermine the appropriacy of its use for L2 learning purposes.

Previously, in Phase 1 of our research, we examined GT transcription accuracy for correctly and incorrectly pronounced items in *predictable* sentence-final contexts (e.g. *I don't know who to thank*$^{\surd}$*-tank*^{X}) (John et al., in press). Correctly pronounced items showed high transcription accuracy rates (88.33%). Real-word output in the error condition showed lower accuracy (47.50%), but considerably higher than nonword output in the error condition (8.33%). GT's ASR is thus particularly good at confirming accurate pronunciation in predictable contexts, especially given that no false alarms were observed (e.g. instances where a correctly pronounced *thank* was transcribed as *tank*, erroneously indicating an error). It also flags pronunciation errors almost half the time, as long as these lead to real words (i.e. *thank→tank* but not *thief→tief*). As summarized in Table 1 below, we thus observed more ('>') confirmative than corrective feedback; and within corrective feedback, more feedback on real than nonwords. Where it failed to transcribe errors accurately, GT's ASR usually produced false negatives (36.66% for real words; 65.00% for nonwords). False negatives are transcriptions that reflect the target item despite incorrect pronunciation (e.g. *thank* mispronounced as *tank* or *thief* mispronounced as *tief* being nonetheless transcribed as *thank* and *thief*). These occur partly because ASR can recover the target item from contextual cues despite incorrect pronunciation. Such cues are, however, exclusively available in predictable contexts, hence the importance of investigating unpredictable contexts. Interestingly, transcription accuracy for female speakers was consistently higher, so the concern that female learners might receive less accurate feedback due to gender bias appears unfounded. Tentatively, we attributed the female advantage to women's generally more targetlike L2 production and careful articulation (Moyer, 2016). This female advantage should be less evident in unpredictable contexts, since clear articulation of the carrier sentence in no way aids identification of the target; indeed, the usual pattern for gender bias, advantaging male speakers, could conceivably emerge in unpredictable contexts.

The current study, constituting Phase 2 of the research, retested GT's ASR for real and nonwords produced in an *unpredictable* carrier sentence. Without contextual cues, the distinction between real words corresponding to correct vs incorrect pronunciations no longer applies. Thus, confirmative and corrective feedback were conflated, and only the *real-word* > *nonword* advantage was investigated ('>' = 'higher transcription accuracy than'). Being identified solely via phonetic cues, decontextualized real words should show lower transcription accuracy than correctly pronounced items (< 88.33%) in predictable contexts (i.e. as observed in our previous study), but higher than incorrectly pronounced items (> 47.50%) in predictable contexts (again, from the previous study). Put differently, we expected GT's ASR to be worse in unpredictable contexts at confirming correct pronunciation, but better at flagging real-word errors (see Table 1 for a summary of these hypotheses). That is, for confirmative feedback, we anticipated a *predictable* > *unpredictable* advantage; whereas for corrective feedback, we anticipated the reverse *unpredictable* > *predictable* advantage. We likewise investigated whether gender bias emerges in unpredictable contexts or whether the *female* > *male* advantage persists, and we

gathered information on false alarms/negatives (which, like confirmative and corrective feedback are necessarily conflated in unpredictable contexts).

Table 1. Summary of Phase 1 findings (predictable contexts) vs Phase 2 hypotheses (predictable contexts)

PHASE 1 (predictable contexts)	PHASE 2 (unpredictable contexts)
Real words: *confirmative* > *corrective feedback* 88.33% vs 47.50%	Real words: *confirmative*/*corrective feedback* < 88.33% *confirmative*/*corrective feedback* > 47.50%
real words > *nonwords* 88.33% / 47.50% vs 8.33%	*real words* > *nonwords*
F > M	M > F or F > M
False alarms: 0%	False alarms: no hypothesis formulated
False negatives: 36.66% (real words); 65.00% (nonwords)	False negatives: no hypothesis formulated

2. Method

Ten Male (M) and 10 Female (F) QF adults were used to record 200 items in a carrier sentence (*This is what I would like to say*, "_____"). The recordings were not based on spontaneous speech with naturally occurring errors and correct pronunciations. Instead, we asked speakers to produce th-initial, t-initial, h-initial, and vowel-initial real and nonwords as presented in a written prompt, and any recordings containing genuine mispronunciations were eliminated. That is, the speakers should be viewed as L2 voice actors used to generate stimuli rather than as participants. The true participant in this research is Google Translate itself.

Based on minimal pairs (e.g. *thank-tank*, *hate-ate*), the 140 real-word items we used were th- vs t-initial and h- vs vowel-initial. The 60 nonword targets comprised t-initial, vowel-initial, and h-initial forms (e.g. *tief*, *_appy*, *hice*). The items thus covered all of the output forms under QF correct or incorrect production of the English 'th' and 'h' sounds. Fewer nonwords were tested than real words mainly because we were confident, based on our previous findings during Phase 1 (John et al., in press), that GT would be unable to transcribe these accurately. Of the 20 recordings of each item in the carrier sentence, we selected ten (5M/5F) to play into GT's ASR. In determining which recordings to retain, those with unclear or erroneous articulation of the target items were eliminated, such that only optimal recordings for our research aims remained. These 2000 recordings were coded for final-item transcription accuracy with the aim of comparing real vs nonword output and M vs F speakers.

Inaccurate transcriptions were further investigated for 'false alarms/negatives'. False alarms/negatives involve real words being transcribed as the minimal pair opposite, such as a *thank* recording being transcribed as *tank* or vice versa. This misleadingly suggests learners have substituted 't' for 'th' (false alarm) or correctly realized 'th' when 't' was in fact substituted (false negative). A nonword transcribed as its real-word counterpart (*tief* transcribed as *thief*) also constitutes a false negative.

3. Results & discussion

Table 2 presents accuracy rates for transcriptions of real words produced by male and female speakers, both separately (M/F) and combined (M + F).

Table 2. Transcription accuracy: real words (e.g. *thank, tick, hold, eat*) in unpredictable contexts (%)

Target items	M	F	M + F
th-initial	60.00	64.50	62.25
t-initial	35.00	34.00	34.50
h-initial	70.00	76.00	73.00
V-initial	74.00	84.50	79.25
Mean	**59.75**	**64.75**	**62.25**

As expected, the overall accuracy rate for male and female voices combined (62.25%) is lower than observed in our previous study for correctly realized items (88.33%) in predictable contexts. Conversely (again, as expected), the rate of 62.25% is higher than our previously observed rate for incorrectly realized items leading to real-word output (47.50%) in predictable contexts. One anomaly is that t-initial real words in our current study inexplicably show lower accuracy (34.50%) than incorrectly realized items, leading to real-word output (47.50%) in predictable contexts. GT's ASR transcription accuracy in unpredictable contexts performs equally well for both correct and incorrect pronunciations constituting real words: without contextual information, only phonetic cues participate in item identification, leading to lower accuracy in confirming correct pronunciation but higher in flagging incorrect pronunciation. Corrective feedback on error is thus more reliable in unpredictable contexts, whereas confirmative feedback on correct pronunciation is less reliable.

Furthermore, we can report that false alarms/negatives are rare among real words (2.25-4.25%), with the minor exception again of t-initial items (14.00%). GT's ASR thus tends not to signal that learners, upon producing a real word, have either mispronounced a correctly realized sound or correctly realized a mispronounced sound. Indeed, many inaccurate transcriptions could be designated 'near accurate' (13.75-27.00%), meaning the transcription, while diverging from the actually realized item, nonetheless accurately reflects the quality of the initial sound. For example, output *thank, hold,* and *tank* transcribed as *think*, *home*, and *take;* while strictly speaking this is inaccurate, they are 'near-accurate' insofar as they correctly indicate how the crucial initial sound was produced.

Interestingly, as observed previously in predictable contexts, transcription accuracy for female recordings of real words is higher than for male recordings across nearly all conditions (Table 2). We anticipated that the female advantage might disappear in unpredictable contexts, since careful pronunciation of the neutral carrier sentence (expected in female L2 speech; Moyer, 2016) provides ASR with no advantage in identifying the final item. Nonetheless, clearer female articulation of just the target itself apparently aids item identification. Table 3 presents accuracy rates for transcriptions of nonwords.

Table 3. Transcription accuracy: nonwords (e.g. *tief, hice, elp*) in unpredictable contexts (%)

Target items	M	F	M + F
t-initial	0.00	0.00	0.00
h-initial	8.00	9.00	8.50
V-initial	0.00	2.00	1.00
Mean	**2.67**	**3.67**	**3.17**

The overall mean for nonword output for male and female voices is glaringly low (3.17%), but this is not surprising given that GT cannot match a nonword to an entry in its lexicon. This finding confirms that GT is essentially incapable of providing corrective feedback on nonword mispronunciations. Indeed, the few instances where transcriptions actually captured phonetic output presumed to constitute nonwords, involved instances where GT was able to identify a proper noun (e.g. *hivy* transcribed as *Hy-Vee*, a grocery store) or to segment the input into smaller units (e.g. *hegos* transcribed as *he goes*). These findings suggest that, to be effective, pronunciation activities should focus on target items resulting in real words if mispronounced. We also observed high rates of false alarms/negatives among nonwords (26.00-40.50%), which only reinforces this implication.

Nonetheless, we should point out that many of the inaccurate transcriptions reassuringly constitute 'near accurate' transcriptions (37.50-59.00%). That is, the realization of the initial sound was frequently reflected in the transcription (e.g. the realization *tief* for target *thief* was transcribed as *teeth*, accurately signaling that 't' was substituted for 'th'). Thus, while real-word output in controlled activities is ideal for generating GT's ASR pronunciation feedback, the technology can still generate partial ('near accurate') corrective feedback on nonwords produced in more open activities such as those involving spontaneous speech.

4. Conclusions

GT's ASR can provide beneficial L2 pronunciation feedback. However, our investigation of QF th-substitution, h-deletion, and h-epenthesis reveals that the accuracy of the feedback is affected by the context in which these pronunciation issues occur. For flagging pronunciation errors (corrective feedback), unpredictable contexts are better; for confirming correct pronunciation (confirmative feedback), predictable contexts are. Moreover, regardless of context, GT's ASR is markedly better at flagging incorrect pronunciations that generate real words (*thank→tank*) than nonwords (*thief→tief*). Pronunciation activities should thus take into consideration both the presence/absence of contextual cues and the lexical status of mispronounced items. We suggest teachers create practice sentences in which mispronunciation of the target sounds results in real words, thus increasing the amount of ASR corrective feedback learners receive. Target words could initially be placed in decontextualized carrier sentences or word lists to generate more corrective feedback on pronunciation errors. As students accurately produce the sound more frequently, teachers could employ sentences that use the word in context. Doing so will provide students with more confirmative feedback from ASR, showing students that they can successfully produce the sound and increasing their self-efficacy in their pronunciation skills. It is in our plans to develop and trial activities based on these findings and suggestions. Reassuringly, the anticipated gender bias failed to emerge: even for items in unpredictable contexts, female speakers showed higher transcription accuracy than males. In sum, GT's ASR has considerable potential to generate invaluable feedback on pronunciation, but its ability to provide both corrective and confirmative feedback is influenced crucially by the nature of the sentence context (predictable vs unpredictable) and by the lexical status of output (real word vs nonword).

Acknowledgements

We are grateful to Andrew Conroy for splicing and editing of the sound files. This project received funding from the *Luc Maurice Foundation* and the *Social Sciences and Humanities Research Council* (430-2022-00512).

References

Ashwell, T., & Elam, J. (2017). How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production? *The JALT CALL Journal, 13*(1), 59-76. https://doi.org/10.29140/jaltcall.v13n1.212

Garnerin, M., Rossato, S, & Besacier, L. (2019). Gender representation in French broadcast corpora and its impact on ASR performance. In *AI4TV '19: Proceedings of the 1st international workshop on AI for smart TV content production* (pp. 3-9). Association for Computing Machinery. https://doi.org/10.1145/3347449.3357480

Inceoglu, S., Chen, W., & Lim, H. (2022). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL, 35*(1), 89-104. https://doi.org/10.1017/S0958344022000192

John, P., Cardoso, W., & Johnson, C. (in press). Automatic speech recognition as a source of corrective feedback on L2 pronunciation. In M. Peterson & N. Jabbari (Eds.), *Frontiers in computer assisted language learning* (pp. 1-19). Routledge.

John, P., & Frasnelli, J. (2022). On the lexical source of variable L2 phoneme production. *The Mental Lexicon, 17*(2), 239-276. https://doi.org/10.1075/ml.22002.joh

Moyer, A. (2016). The puzzle of gender effects in L2 phonology. *Journal of Second Language Pronunciation, 2*(1), 8-28. https://doi.org/10.1075/jslp.2.1.01moy

Saito, K. (2021). Effects of corrective feedback on second language pronunciation development. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 407-428). Cambridge University Press. https://doi.org/10.1017/9781108589789.020

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16917

ChatGPT and language learning: University educators' initial response

Louise Ohashi[a] and Antonie Alm[b]
[a]Department of English Language and Cultures, Gakushuin University, Japan, , ohashigakushuin@gmail.com and [b]Languages and Cultures Programme, University of Otago, New Zealand, , antonie.alm@otago.ac.nz

How to cite: Ohashi, L.; Alm, A. (2023). ChatGPT and language learning: University educators' initial response. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16917

Abstract

This article examines the initial response of university-level language educators to ChatGPT, an AI chatbot, within the first ten weeks of its release. ChatGPT, which stands for 'Chat Generative Pre-trained Transformer', is a conversational AI language model developed by OpenAI, a private artificial intelligence research organisation. A global survey collected data from 367 teachers of 16 languages. The results show strong awareness of ChatGPT among educators, but wide variation in experience and knowledge highlights the need for training and support. While use of ChatGPT was limited, particularly for pedagogical tasks, the majority of educators expressed interest in integrating it into their future teaching practice. In more concrete terms, more teachers indicated they were likely to use it to create language learning resources and recommend it for self-study than use it for managing assessment. In regards to the potential impact of ChatGPT on the field of education, educators showed a mixture of concern and optimism. While many agreed that it would increase the accessibility of language teaching and have a positive impact on self-study, even more had concerns about the potential for cheating and over-reliance, highlighting the importance of ethical guidelines. This research contributes to the understanding of the opportunities and challenges associated with ChatGPT in language education.

Keywords: *ChatGPT, artificial intelligence, language education*

1. Introduction

ChatGPT, a conversational AI language model with wide-ranging capabilities, has taken the world by storm since its release on November 30, 2022. Chatbot technology is not new but its speed of adoption certainly is, with Reuters reporting an estimated 100 million monthly active users within two months of release, "making it the fastest-growing consumer application in history" (Hu, 2023, February 3). High adoption rates at a societal level meant that educators needed to respond soon after its release, and within seven weeks a *New York Times* headline had already proclaimed "universities start revamping how they teach" (Huang, 2023, January 17). However, the extent to which this was happening and how educators felt about it was unclear. To investigate the initial response to ChatGPT within language education, the researchers opened a survey to foreign language educators at universities worldwide. Three research questions guided the study:

RQ1. To what extent had language educators heard about and used ChatGPT within ten weeks of its release?

RQ2. What were language educators' views on adopting ChatGPT for educational purposes within ten weeks of its release?

RQ3. What were language educators' perceptions of the potential impact of ChatGPT on language education within ten weeks of its release?

2. Method

2.1. Instrument

To address the research questions, the authors created a survey based on their own beliefs about possible areas of importance and output from ChatGPT about its possible roles in language education. Their approach included prompting ChatGPT for output on relevant issues, discussing the output, creating Likert scale items with points deemed relevant, prompting for more information, and refining the items after further discussion. The survey collected background information about participants and data on three key areas: 1) teachers' knowledge about and use of ChatGPT; 2) current and likely use of ChatGPT to create language learning resources, manage assessment, and advise students for self-study; and 3) views on ChatGPT's impact on language education. The survey combined open and closed questions with Likert scale items most commonly employed. Before its distribution, the survey was piloted by four colleagues of the researchers, with modifications made to improve readability.

2.2. Context and participants

The survey was distributed via email, Facebook, and Twitter, with encouragement from the researchers for others to share it widely. Snowball sampling was used to reach far beyond the researchers' personal networks, a method promoted by Leighton et al. (2021). The survey was open for less than two weeks and closed exactly ten weeks after ChatGPT's release to capture an initial response to its impact. Valid responses were received from 367 anonymous participants (51.5% female, 44.7% male, 0.5% non-binary, 3.3% undisclosed) who taught 16 languages[1] at universities around the world (48 countries/regions[2]).

2.3. Data analysis

Descriptive statistics were calculated for quantitative items. Qualitative data underwent thematic analysis. Key quantitative findings are reported in the results section. Due to space limitations, detailed statistical analyses and qualitative themes are not presented here but will be disseminated through future publications and conference presentations.

3. Results and Discussion

3.1 ChatGPT knowledge and experience (RQ1)

Within ten weeks of ChatGPT's release, 22.3% of respondents already claimed to have extensive knowledge of it ("I have heard the name and know a lot about it"), and 62.7% had heard of it and had some understanding. In contrast, only 15% had very limited or no knowledge (7.9% had heard of it but knew nothing about it, 7.1% hadn't heard of it). This distribution shows strong awareness of ChatGPT among language educators. Usage was almost evenly split (50.1% had used it, 49.9% hadn't). This suggests that although most language teachers were aware of ChatGPT, only half had taken the next step to usage.

The overall mean score for familiarity with ChatGPT was 4.31 on a scale of 0 to 10, indicating a relatively low level of familiarity. However, 27% chose 7 or higher, suggesting that while the average familiarity may be low, a substantial subset of respondents were more familiar with ChatGPT.

To gain insight into educators' perceptions of their ability to use ChatGPT for educational purposes, they were asked how well they felt they could carry out specific tasks based on their current level of familiarity with the tool. The results (Table 1) showed that few had high confidence in their ability to create language learning

[1] Languages participants taught (alphabetical order): Ancient Greek, Chinese, English, French, German, Italian, Japanese, Latin, Madurese, Norwegian, Portuguese, Russian, Spanish, Swedish, Thai, and Welsh.

[2] The country list built into Qualtrics listed some countries together as regions (such as the UK). This was noticed during analysis.

resources (36.8% extremely/very well), guide students towards self-study (28.6% extremely/very well), or create automated assessment tasks (20.1% extremely/very well), and 23.7%-38.2% noted they would not be able to do these tasks well at all.

Table 1. Language educators' beliefs about their ability to do teaching-related tasks with ChatGPT

	Extremely well (5)	Very well (4)	Moderately well (3)	Slightly well (2)	Not at all well (1)	*M*	*SD*
Create language learning resources	13.6%	23.2%	24.0%	15.5%	23.7%	2.87	1.37
Create automated assessment tasks	8.7%	11.4%	24.0%	17.7%	38.2%	2.35	1.32
Guide students on self-study tasks	10.1%	18.5%	27.3%	19.3%	24.8%	2.70	1.30

A significant majority of language educators had not incorporated ChatGPT into their teaching practice, at least not in the ways the researchers targeted. Specifically, only 20.2% had used it to create language learning resources, 6.5% had used it for assessment purposes, and 13.1% had advised students to use it for self-study.

In summary, the results indicate that within the first ten weeks of its release, there was widespread awareness of ChatGPT among language teachers, but limited use.

3.2. Views on potential use of ChatGPT in language education contexts (RQ2)

Although adoption of ChatGPT was uncommon, a significant proportion of teachers expressed interest in integrating it into their teaching, reflecting its potential value in language education. Specifically, 40.9% were very interested, 27.5% were interested, and 26.4% were unsure. Only a small minority showed disinterest (3.3% not interested, 1.9% not interested at all). The mean and median scores were both 4 (SD = 0.989) on a scale of 1 to 5, indicating strong overall interest in using ChatGPT in future classroom activities.

The survey noted, "ChatGPT claims it can be used to create the following language learning resources. How likely are you to use it for [various tasks]?" (Table 2). More than half of the participants responded positively (extremely/somewhat likely) about potential use for creating language learning resources, particularly for improving reading (66.7%) and writing (62.6%) skills. With extra time to explore its potential, interest may have turned into action for more teachers, as reported in a recent study in Vietnam (Nguyen, 2023) that found English teachers made extensive use of ChatGPT for lesson planning and material creation.

Table 2. Language educators' likelihood to use ChatGPT to create learning resources

	Extremely likely (5)	Somewhat likely (4)	Neither (3)	Somewhat unlikely (2)	Extremely unlikely (1)	*M*	*SD*
Vocabulary: Generate flashcards with vocabulary words and their definitions, translations or synonyms.	16.9%	36.0%	18.8%	14.4%	13.9%	3.28	1.29
Grammar: Generate worksheets with grammar rules, examples, and exercises.	16.6%	37.1%	19.3%	13.4%	13.6%	3.30	1.28
Writing: Generate templates for different types of writing, such as an essay, a letter, or a report.	26.4%	36.2%	18.3%	12.3%	6.8%	3.63	1.19

Reading: Generate reading passages or texts on different topics and levels of difficulty.	27.5%	39.2%	16.1%	9.0%	8.2%	3.69	1.12
Conversation: Generate scripts for role-playing activities.	20.7%	34.3%	17.2%	13.4%	14.4%	3.34	1.33
Quizzes and tests: Generate quizzes and tests on different language skills, such as vocabulary, grammar, reading comprehension, and listening.	19.1%	34.6%	21.3%	11.1%	13.9%	3.34	1.29
Games: Generate language learning games such as crossword puzzles, word searches, and hangman games.	20.4%	33.2%	20.2%	10.9%	15.3%	3.33	1.33

In terms of assessment, teachers were more likely to envisage using ChatGPT for automated scoring of multiple-choice tests (49.8%) and automated feedback on writing tasks (42.5%) than automated scoring of writing tasks (34.3%). When advising on self-study tasks, they were most likely to encourage students to use ChatGPT as a virtual conversation partner (59.7%) and a virtual tutor (56.9%). In addition, they viewed ChatGPT as a potential tool for creating engaging learning environments, with 56.1% likely to advise students to use ChatGPT to generate interactive language learning games.

These data suggest a marked contrast between the perceived potential of ChatGPT and its actual use in classroom practice during the first ten weeks after its release. While the majority of respondents recognised the benefits of ChatGPT for resource creation, assessment, and self-study, a significantly smaller percentage had implemented it in these areas. This trend suggests a gap that could be filled with further research, training, and guidance on how to use ChatGPT effectively in language teaching contexts.

3.3. Perceptions of the potential impact of ChatGPT on language education (RQ3)

Respondents showed a mixture of optimism and concern when assessing the potential impact of ChatGPT on language teaching, as shown in Table 3. In terms of positive sentiments, 58% of respondents believed that ChatGPT would lead to an improvement in the accessibility of language teaching and 55.1% expected it to have a positive impact on self-directed learning. Furthermore, 53.7% believed that ChatGPT could lead to more personalised language learning.

However, these positive sentiments were accompanied by notable concerns. Specifically, 70.8% of respondents agreed that the introduction of ChatGPT could lead to an increase in cheating and academic dishonesty. In addition, 62.2% were concerned that students would become overly reliant on ChatGPT, thereby inhibiting the development of their language and critical thinking skills.

There were more measured responses to other potential impacts. For example, only 37.1% of respondents believed that ChatGPT would lead to more efficient and effective language assessment. Meanwhile, almost half (48.5%) agreed that the introduction of ChatGPT could reduce reliance on human teachers for certain aspects of language teaching and possibly reduce the cultural and human aspects of language education.

Table 3. Educators' level of agreement on the potential impact of ChatGPT on language education

	Strongly agree (5)	Somewhat agree (4)	Neither (3)	Somewhat disagree (2)	Strongly disagree (1)	*M*	*SD*
More efficient and effective language assessment	10.9%	26.2%	41.7%	13.6%	7.6%	3.19	1.05
More personalized language learning	17.2%	36.5%	30.2%	9.3%	6.8%	3.48	1.09
More self-directed language learning	15.3%	39.8%	32.1%	7.9%	4.9%	3.53	1.00
Reduced dependence on human teachers for certain aspects of language instruction	13.9%	34.6%	28.6%	16.1%	6.8%	3.33	1.11
An increase in cheating and academic dishonesty	35.4%	35.4%	22.9%	4.4%	1.9%	3.98	0.97
A decrease in the cultural and human aspects of language education	14.7%	33.8%	29.4%	16.9%	5.2%	3.36	1.09
Students will become too reliant and not develop their own language skills and critical thinking abilities	24.3%	37.9%	24.5%	10.4%	3.0%	3.7	1.04
An improvement in the accessibility of language education	17.7%	40.3%	30.5%	6.8%	4.6%	3.6	1.00

Looking more broadly, 52.6% of respondents saw the potential for ChatGPT's overall impact on the field of language education to be positive (12% very positive, 40.6% somewhat positive). Educators mainly envisaged benefits in terms of accessibility, self-directed learning, and personalisation, echoing recent findings on positive teacher and student attitudes towards ChatGPT's potential for independent learning (Ali et al., 2023). However, participants in the present study also expressed concerns about an increased potential for academic dishonesty and over-reliance, as well as a potential reduction in the human-centred aspects of language teaching, signalling areas that educators need to proactively address.

4. Conclusions

This study provides insights into language educators' experiences and perceptions of the role and potential of ChatGPT in language education, collected within ten weeks of its release. The results show an overall high level of awareness of ChatGPT and considerable interest in using it for educational purposes. However, usage in classroom practice was less widespread, suggesting a contrast between its perceived potential and actual use. Perceptions of its impact on language education showed educators identified both opportunities and challenges, which suggests there will be much need for guidance in navigating paths towards pedagogical use and overcoming potential problems. While ChatGPT is seen as a means to improve the accessibility of language teaching and encourage self-directed learning, it also raises concerns about academic dishonesty, over-reliance on technology and a potential reduction in the human and cultural elements of language education.

The authors are currently carrying out a more detailed analysis of the quantitative and qualitative data from the survey. This will allow a deeper exploration of the complex and nuanced experiences, perspectives, and challenges language teachers face when incorporating ChatGPT into their practice. It is anticipated that this will

provide additional insights into the implications and potential applications of ChatGPT in language teaching, further enriching the knowledge and discourse surrounding the use of AI in education.

Acknowledgements

We would like to thank all of our colleagues around the world—those who know us and those who do not— who helped distribute the survey and shared their valuable feedback. We would also like to thank this project's research assistant, Pei Jun See, for her help with data cleaning and preparation for analysis in SPSS.

References

Ali, J., Shamsan, M., A., A., Hezam, T. & Mohammed A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Hu, K. (2023, February 3). ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Huang, K. (2023, January 16). Alarmed by A.I. chatbots, universities start revamping how they teach. *The New York Times*. https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html

Leighton, K., Kardong-Edgren, S., Schneidereith, T., & Foisy-Doll, C. (2021). Using social media and snowball sampling as an alternative recruitment strategy for research. *Clinical Simulation in Nursing, 55*, 37-42. https://doi.org/10.1016/j.ecns.2021.03.006

Nguyen, T. T. H. (2023). EFL teachers' perspectives toward the Use of ChatGPT in writing Classes: A case study at Van Lang University. *International Journal of Language Instruction, 2*(3), 1-47. https://doi.org/10.54855/ijli.23231

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16993

Exploring users' perceptions of ASR for writing narrative texts

Danial Mehdipour-Kolour[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, Montreal, danial.mehdipourkolour@mail.concordia.ca and [b]Department of Education, Concordia University, Montreal, walcir.cardoso@concordia.ca

How to cite: Mehdipour-Kolour, D.; Cardoso, W. (2023). Exploring users' perceptions of ASR for writing narrative texts. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16993

Abstract

This study examines how users perceive Automatic Speech Recognition (ASR) as a tool for writing narrative texts, and compares the perceptions of two groups of users: native and non-native English writers. As such, this study aimes to answer the following questions: (1) How do English writers perceive the use of ASR as a writing tool?; and (2) How do native and non-native English writers' perceptions compare in terms of using ASR as a writing tool? To answer these questions, we employed the Technology Acceptance Model 2 (TAM2) to investigate 60 participants' perceptions of utilizing ASR for producing narrative texts. Our findings from analyzing seven components of TAM2 show that writers express a positive attitude towards utilizing ASR as a tool for composing texts. Our findings also indicate no noticeable differences between how native and non-native English writers perceive the usefulness of ASR for creating texts. This is contrary to our hypothesis that native speakers, owing to their more advanced pronunciation skills in English, might have a more favorable attitude towards using ASR.

Keywords: *automatic speech recognition, L2 writing, technology acceptance model, user perception.*

1. Introduction

This study examines the use of ASR as a writing tool and, more specifically, how writiers perceive its use for composing narrative texts, a genre that tells a story in the form of a novel or personal essay. Narrative genre is often considered "cognitively easier and [uses] less complex language than […] exposition and argumentation" (Weigle, 2010, p. 100). ASR identifies voice input and converts it to text output, which is automatically displayed on the screen. Although the utilization of voice as an *input medium* is increasing, particularly in mobile settings (Sohail, 2020), there remains limited knowledge about users' perceptions of using ASR as a writing tool, especially considering new and more advanced speech-based technologies, such as those found in Google Voice Typing.

While research has demonstrated that ASR is useful for enhancing second language (L2) oral skills, such as pronunciation and speaking (Cox & Davies, 2012; Liakin et al., 2015), there is no existing evidence of how adult native and non-native English writers without learning difficulties perceive the use of ASR for writing texts. Moreover, no research has investigated whether there are differences between these two groups of writers in perceiving the usefulness of this technology for writing narrative texts. In fact, excluding a recent study by Johnson and Cardoso (in press) involving only L2 writers, research of this type has not been conducted for approximately two decades (e.g. Leijten & Van Waes, 2005; MacArthur, 1999). In addition, apart from Johnson and Cardoso (in

press), the limited existing studies have examined ASR solely with young learners, e.g. to assist literacy development or writers with learning difficulties (Quinlan, 2004).

Following Cardoso's (2022) chronological framework for conducting CALL research, this study focuses on an early stage of investigation by examining users' perceptions of ASR as a writing tool. As Venkatesh and Davis (2000) indicate, analyzing users' perceptions provides insights into factors influencing acceptance and potential adoption of a target technology. According to the authors, *perceived usefulness* and *ease of use* are key determinants of users' intention to adopt a new technology. To examine users' perceptions, we employed Venkatesh and Davis' (2000) Technology Acceptance Model 2 (TAM2), which is implemented in the form of a survey with a 5-point Likert scale (ranging from 1 = strongly agree to 5 = strongly disagree). The survey included the following seven constructs: (1) Perceived Ease of Use (PEU) (or the extent the user believes that using the technology will be effortless; e.g. *I find ASR easy to use*); (2) Perceived Usefulness (PU) (or the extent the user believes that using the technology will improve their performance; e.g. *Using ASR improves my writing performance*); (3) Intention to Use (IU) (e.g. *I intend to use ASR as a writing tool*); (4) attitude (AT) (e.g. It is a good idea to use ASR for writing purposes); (5) Subjective Norm (SN) (or the perception that is shaped by the influence of external stakeholders like school administrators; e.g. *As a university student, it is important for me to use ASR for writing purposes*); (6) Output Quality (OQ) (or the extent to which the technology can perform the target tasks effectively; e.g. *The quality of ASR-produced text is high*); and (7) Result Demonstrability (RD) (or the extent the results of using the technology are demonstrable; e.g. *I believe I can share my experience using ASR with others*). Each construct of the survey serves as a predictor for the user's perception and acceptance of ASR as a tool for writing.

Given that speech recognition systems, often trained on native speech, may struggle with non-native accents (Liakin et al., 2015), this study moreover compared native and non-native English writers' perceptions of using ASR as a tool for composing narrative texts.

Based on the above, the study was guided by the following Research Questions (RQs):

(1) How do users perceive the adoption of ASR as a tool for producing narrative texts?

(2) How do native and non-native English speakers' perceptions differ regarding the use of ASR as a tool for writing narrative texts?

2. Method

2.1. Context and participants

This study recruited 60 participants, including Native Speakers (NSs; *n*=30) and Non-Native Speakers (NNSs; *n*=30), between the ages of 20 and 50, with a gender distribution of 22 males and 37 females. Participants were undergraduate and graduate students at an Anglophone university in Montreal, Canada. As is customary in research involving humans, consent was obtained from all participants, who were compensated $20 for their participation. They were recruited using social media and other calls for participation, such as email lists. Participants spoke a variety of native languages (e.g. Farsi, French). Due to the COVID-19 pandemic, this study was carried out online using the Zoom video conferencing platform, in individual sessions in which each participant interacted with the ASR for writing their narrative texts. These sessions were video-recorded for analysis (reported in another study).

2.2. Procedures and instruments

Participants were asked to use Google Voice Typing (Google's ASR engine) to produce texts in Google Docs, the application selected for this research. Firstly, participants were assigned free writing activities (e.g. a visual writing prompt) to familiarize themselves with ASR-based writing and to promote fluency in text production (Tynjälä et al., 2001). This phase was followed by a dictation activity, including two short passages. This activity prepared participants for writing longer texts with ASR and helped participants to adjust their writing behavior to punctuate and format their texts, thus becoming more acquainted with ASR features and its writing potential. After practicing

writing with ASR, participants were asked to complete the narrative writing section. Three different narrative prompts were presented (e.g. *success story: write a story about something you accomplished*), and participants were instructed to select two of them and, within 20 minutes, write their narratives using ASR. After the successful completion of the assigned tasks, participants were requested to take part in a 20-minute survey informed by the TAM2 model. The survey was administered online using Google Forms (see sample questions in Section 1), a well-known and reliable platform for performing surveys.

2.3. Data analysis

To find out how users perceive the adoption of ASR as a tool for writing narrative texts (RQ1), a one-sample t-test was administered. Prior to conducting the t-test, a Shapiro-Wilk test was run to assess the normality assumption. The test conducted evaluated whether participants' perception scores were statistically different from *neutral*, which was considered the population perception..

3. Results

Shapiro-Wilk's test revealed that perception scores were normally distributed (p >.05) with no outliers detected. Tables 1 and 2 show the results of the one-sample t-test and descriptive statistics for each unique perception-informed construct examined. It can be asserted that, in general, the participants developed a neutral opinion towards their use of ASR as a writing tool (the lower the Means, the higher the acceptance), even though these results demonstrate the participants' acceptance of several of the TAM2 constructs (i.e. PEU, AT, and RD).

Table 1. Descriptive statistics.

Statistics	**PEU**	**PU**	**IU**	**AT**	**SN**	**OQ**	**RD**
M	2.20	3.08	3.12	2.40	3.87	2.95	1.94
Mdn	2.16	3.00	3.12	2.33	4.00	3.00	1.75
SD	.72	.90	1.05	.89	.88	.84	.97
R	3	3.83	4	4	3	3.67	4

Table 2. One-sample t-test results.

Variable	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**
PEU	-6.399	59	.000	-.59722
PU	2.479	59	.016	.28889
IU	2.377	59	.021	.32500
AT	-3.453	59	.001	-.40000
SN	9.450	59	.000	1.07500
OQ	1.434	59	.157	.15556
RD	-6.791	59	.000	-.85833

To address the second research question, which explored the differences in perceptions between native and non-native English speakers regarding their use of ASR as a writing tool, a Mann-Whitney U test was conducted. The findings obtained indicate that there was no statistically significant difference between native and non-native English speakers' perceptions across most TAM2 constructs adopted in this study. However, it was observed that native speakers exhibited a higher mean than their counterparts in PEU (see Table 3).

Table 3. Independent-samples Mann-Whitney U test tesults: Comparing native and non-native writers.

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

4. Discussion

This study examined how users perceive ASR as a tool for writing narrative texts (RQ1) and whether there is a difference in perception among two groups of users: native and non-native English writers (RQ2). Regarding the first question, the results reported indicate that participants' perceptions were slightly more positive compared to average population norms. This suggests that the majority of ASR writers who participated in this research either agreed with or had a neutral stance concerning the utilization of ASR as a tool for composing texts. The presence of such a view can be explained by the user-friendly interface of ASR (van Doremalen et al., 2016). In addition, the use of one's speech fosters fluency (and possibly efficiency) in writing, since it eliminates or minimizes transcription problems, thus alleviating the writer's cognitive burden in the writing process (MacArthur & Cavalier, 2004). However, the results also revealed issues with the ASR tool (observed during the treatment), which may have influenced the participants' sometimes lukewarm responses. Specifically, the adopted ASR system sometimes struggled to accurately recognize certain words when spoken rapidly or inacurately. Additionally, the technology lacked proper text formatting and punctuation without providing verbal cues like "full stop" or "question mark" at the end of each utterance.

For the second research question, our findings showed that native and non-native English writers had comparable views on using ASR as a writing tool. This aligns with recent research indicating that ASR output quality is not substantially impacted by accented speech (see McCrocklin & Edalatishams, 2020), due to recent developments in speech recognition. Therefore, non-native speakers could potentially gain similar benefits from ASR technology as native speakers, due to ASR's ability to accurately transcribe L2-accented speech.

5. Conclusions

The aim of this study was to explore users' perceptions of employing ASR as a tool for writing narrative texts and to find out whether there is a different in perception among native and non-native English writers. The findings suggest ASR engines like Google Voice Typing are beneficial writing aids for both native and non-native English speakers. One possible factor influencing positive user perceptions of ASR-assisted writing is the technology's

ability to accurately transcribe both native and accented speech. These results make important contributions to the literature in two key areas. First, we examined user perspectives on employing ASR for writing, providing valuable insights into this relatively unexplored topic. Second, we focused on comparing native and non-native English speakers, an important yet overlooked subject in prior ASR writing studies. Overall, this work expands current knowledge by eliciting users' views on ASR-facilitated writing while highlighting differences based on pronunciation proficiency of native and non-native speakers of English.

References

Cardoso, W. (2022). Technology for Speaking Development. In T. Derwing, M. Munro, & R. Thomson (Eds), *Routledge Handbook on Second Language Acquisition and Speaking* (p. 299-313). Routledge, Taylor & Francis. https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Speaking/Derwing-Munro Thomson/p/book/9780367900847

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

Johnson, C. & Cardoso, W. (in press). Hey Google, let's write: Examining L2 learners' acceptance of automatic speech recognition as a writing tool. *CALICO Journal*.

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

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

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

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

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

Perreault, C., & Mathew, S. (2012). Dating the origin of language using phonemic diversity. *PloS One*, *7*(4), 1–8.

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

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

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

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

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

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

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

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16919

Machine translation in language education: Perspectives from advanced language learners

Antonie Alm[a] and Yuki Watanabe[b]

[a]Languages & Cultures Programme, University of Otago, , antonie.alm@otago.ac.nz and [b]Media, Film and Communication Studies Programme, University of Otago, , yuki.watanabe@otago.ac.nz

How to cite: Alm, A.; Watanabe, Y. (2023). Machine translation in language education: Perspectives from advanced language learners. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16919

Abstract

This paper explores the experiences and perspectives of advanced language learners on the use of Machine Translation Tools (MTTs). The study focuses on how learners navigate their use of tools like Google Translate, DeepL, and Reverso, and how their usage is influenced by instructional recommendations. Nine learners, studying different languages at an advanced level, were interviewed in focus groups. Thematic analysis of the interview transcripts revealed key themes related to the use of Online Language Learning Tools (OLLTs) and MTTs, students' uncertainty about MTT functionalities, policies regarding MTT use, and suggestions for MTT training. The findings highlight the significant impact of initial encounters with MTTs on learners' long-term perceptions and use of these tools. Participants demonstrated strategic use of MTTs to support their reading, writing, listening, and speaking skills, while being aware of their limitations. Although learners found MTTs helpful, they also emphasised the importance of authentic communication in language learning. The study concludes with recommendation for a collaborative approach between teachers and learners to effectively integrate MTTs into the language learning curriculum.

Keywords: *machine translation, google translate, advanced language learners, role of instructors.*

1. Introduction

Machine Translation Tools (MTTs) have become widespread in language learning, resulting in increased speed, fluency, accuracy, and overall quality of student writing (Briggs, 2018; Chung & Ahn, 2021; Kok Wei, 2021; Kol et al., 2018; Lee, 2020; Tsai, 2019;). However, their effectiveness is still debated among teachers and students (Cowie & Sakui, 2023). Studies show a link between MTT use and learner proficiency, more advanced learners tend to use MTTs for vocabulary searches while low proficiency learners often check grammar (Ahn & Chung, 2020). In addition, higher proficiency learners are more likely to critically evaluate MTT output to ensure it matches their intended meaning, using it as a tool to improve their writing. In contrast, lower proficiency learners tend to more readily accept the unmodified MTT output without further modification, focusing on micro-level errors.(Lee & Briggs, 2021; Chung & Ahn, 2021).

Building on our previous work (Alm & Watanabe, 2022), which revealed proficient MTT use among advanced language learners, this study further explores advanced students' experiences with MTTs. That research found

students valued the convenience and guidance MTTs provided for writing, but also recognised the importance of critically evaluating output and avoiding over-reliance on MTTs.

This study aims to extend our understanding of advanced language learners' experiences of MTTs and their associated learning trajectories. It also explores how this understanding might be used to improve the integration of MTTs into language teaching. The guiding research questions are:

1. How do advanced language learners experience MTTs in their learning process?

2. How can these experiences inform the successful integration of MTTs into language teaching?

2. Method

In September 2022, an email invitation was sent to 64 third-year language students at a university in New Zealand to participate in an interview on their experiences with MTTs. Nine students responded and participated, divided into four focus groups with 2-3 participants each. The groups included students of French, German, Japanese, and Spanish at B2 level, aged 19-30.

The interviews, conducted in English by one of the researchers, lasted approximately one hour. They were audio recorded and transcribed in real time using Otter AI. To protect the anonymity of the participants, pseudonyms were used to report the findings. Ethical approval was obtained for this study and informed consent was received from all participants prior to commencement of the interviews.

The two researchers independently coded the transcripts then discussed and agreed on final coding. Inductive thematic analysis was conducted following the six phases outlined by Braun and Clarke (2006) to identify recurring patterns relevant to the research questions. An inductive, data-driven approach was taken, with themes derived from the participant data rather than imposed by pre-existing concepts or theories (Lew et al., 2018). Braun and Clarke's framework guided the coding, collating, reviewing, defining, and narrating of themes from the qualitative data through an iterative, recursive process.

3. Findings

3.1. Use and preference of Machine Translation Tools (MTTs) and Online Language Learning Tools (OLLTs)

The participants revealed that they used different MTTs and OLLTs to support their language learning process. Among the tools used, such as DeepL, Reverso, Google Translate, and language-specific ones like Leo Dictionary, WordReference, SpanishDict, and Takoboto, preferences varied among the participants. For example, Cathy preferred Reverso for its contextual feature, as she explained: "it'll translate like a word or a small phrase for you and then it'll show you a bunch of examples of like how it's been used on the internet … like one or two sentences from like a news story or a blog post … I find it really valuable to see it in like actual French in actual context." Ash on the other hand appreciated DeepL for providing alternative expressions, "it gives you options of like just a few different ways you could say sort of the same thing." Google Translate was mentioned for its usefulness in quickly looking up unfamiliar words or phrases and for travelling. However, it was seen as less suitable for academic learning due to scepticism about its accuracy and naturalness. These advanced students recognised the limitations of MTTs and balanced their use with language specific OLLTs mentioned earlier, which they found effective in improving grammar knowledge, increasing vocabulary, and facilitating a more nuanced understanding of context and word usage.

3.2. Proficiency in using MTTs among advanced learners

The participants reported insights in their use of MTTs for a range of learning activities. They demonstrated proficiency in using MTTs to pre-read challenging texts, to proofread their writing, to check for specific words

or phrases, and occasionally to check pronunciation. For example, Beth and Judy used Google Translate's camera translation feature to quickly understand paragraphs during pre-reading activities. This approach was particularly useful for grasping the main ideas of difficult passages when reading novels.

Learners also reported resourcefulness in their writing process. As well as using MTTs to look up words and phrases before writing, they found reverse translation useful for proofreading. This technique involves translating their work from the target language back to their first language to identify areas for improvement. For example, Sam reported that she found Google Translate extremely helpful in ensuring gender agreement of nouns and adjectives in her French writing.

In terms of listening and speaking, many participants used MTTs to check specific words or phrases and, despite recognising the limitations of automated voices, to hear the pronunciation of individual words. These strategies reflect an overarching skill in the use of MTTs among advanced learners. As suggested by Cathy, "[at the advanced level] everyone knows how to use the tools they like," indicating a shared understanding among their peers of using these tools effectively.

3.3. Evolution of MTT usage from high school to university

The use of MTTs among the learners evolved from an initial reluctant reliance in high school to a more strategic and productive approach in university. Some participants reported to have been discouraged from using MTTs by their high school teachers. Teachers stressed the importance of understanding grammar and being able to formulate sentences independently, and even explicitly condemned the use of Google Translate. Instead, they recommended alternative tools such as Reverso, WordReference, and Leo Dictionary as reliable online dictionaries. Cathy, for example, recalled that her high school French teacher warned her against using Google Translate, saying "it's terrible. Don't ruin your French!" and suggested Reverso, which Cathy still uses today.

Despite the general discouragement from teachers, students found MTTs useful for home assessments because of their accessibility. This held particularly true during the periods of lockdown amidst the COVID pandemic, which often meant a lack of consistent monitoring. Some participants, such as Kate, admitted using Google Translate out of necessity, explaining that she did not "know anyone who didn't cheat their way through those writing assessments." Students' experiences in high school therefore shaped their initial perceptions and use of MTTs, which were heavily influenced by teachers' suggestions and the assessment format.

At university, participants' understanding and use of MTTs evolved. They found that teachers' recommendations often had a positive impact on their language learning. Ash recalled her German instructor introducing DeepL, which she found valuable for translating and understanding texts. Similarly, Kate's curiosity was piqued when her French teacher used Reverso for a translation, which led to her own exploration of Reverso. As the students' learning progressed, they acknowledged that the quality f the MTTs had improved, mentioning Google Translate in particular. Although the use of MTTs was not explicitly discouraged at university level, assessments were designed with MTTs in mind to prevent learners from 'cheating' through the use of MTTs. For example, students were sometimes required to mark where they had used MTTs in their assignments. This practice, together with class discussions about independent language practice (including the use of MTTs), seemed to have fostered an understanding of the potential of MTTs for language learning. Amy summed up this feeling by saying, "if you actually want to learn, [MTT]'s going to provide you with more resources than you can use," acknowledging the value of MTTs when used for active learning.

3.4. Expectations and suggestions from advanced learners

Despite their proficient use of MTTs, participants expressed uncertainty about their teachers' perceptions and acceptance of the use of MTTs in the language classroom. Ash remarked on the negative attitude of some teachers towards MTTs. This was echoed by Cathy and Judy, who emphasised the need for teachers to

acknowledge the existence and recognise the potential of MTTs. Beth noted inconsistent attitudes among her teachers and expressed concern about the lack of open discussion about MTTs in the classroom. As Judy states: "Tell us how you feel about it and what you're comfortable with us doing." Participants strongly recommended clear guidelines and expectations for MTT use to reduce student confusion. Cathy advocated an informative approach to guide students towards responsible and independent use of MTTs for language learning by including an MTT policy in course outlines and providing MTT training to effectively guide students. Fiona suggested that teachers should have open discussions with students about the appropriate use of MTT and its benefits. Ash suggested creating opportunities for students to share their innovative use of MTTs in small group discussions.

Participants recognised the usefulness of MTTs in certain scenarios, however, they also emphasised the importance of authentic communication and understanding cultural nuances. They advocated challenging monolingualism and actively engaging with other languages. Ash emphasised the joy of authentic communication: "I can't think of anything else that gives you the same feeling of achievement when you can communicate with other people." Sam stressed the importance of making an effort to learn a language and challenging the assumption that everyone speaks English. Participants' clear goals and strong motivation seemed to drive their productive use of MTTs.

4. Discussion

This study provides valuable insights into advanced language learners' experiences and perceptions of the use of MTTs. It highlights the influence of students' initial exposure to MTTs on their perceptions and use of these tools. In this sense, educators have a crucial role to play in shaping these attitudes and should be aware of the lasting impacts their guidance can make on learners. To encourage critical use of MTTs, and to alleviate the widespread student anxiety and uncertainty identified by Cowie and Sakui (2023), educators need to provide explicit guidance. Careful consideration should also be given to assessment formats, with activities designed to encourage the effective use of MTTs while promoting independent language comprehension and production. The emergence of advanced language models such as ChatGPT adds an interesting dimension to the conversation. In line with the findings of this study, learners with clear goals and strong motivation are more likely to use MTTs productively. Despite recognising the limitations of MTTs, advanced learners used them as effective support tools, adopting strategies such as pre-reading and reverse translation for different learning contexts.

5. Conclusions

This study explored the perspectives and experiences of advanced language learners regarding the use of MTTs in their learning journey. The findings highlight the importance of the initial encounter with MTTs in shaping students' perceptions and use of these tools over time. From hesitant reliance in high school to strategic use at university, it is clear that the role of the teacher in guiding learners is crucial.

The potential of MTTs, despite their limitations, is recognised by learners. They valued these tools for specific purposes, such as pre-reading challenging texts or proofreading written work. However, the study also emphasised that authentic communication and understanding cultural nuances remained a priority for these advanced learners, highlighting the importance of a balanced approach to the use of MTTs.

The study findings suggest that teachers collaborate with students to clarify the effective use of MTTs, integrate them into language learning tasks, and emphasise the importance of real communication beyond machine translation in language teaching.

References

Ahn, S., & Chung, E. S. (2020). Students' perceptions of the use of online machine translation in L2 writing. *Multimedia-assisted language learning*, *23*(2). https://doi.org/10.15702/mall.2020.23.2.10

Alm, A., & Watanabe, Y. (2022). Online machine translation for L2 writing across languages and proficiency levels. *Australian journal of applied linguistics*, *5*(3), 135-157. https://doi.org/10.29140/ajal.v5n3.53si3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Briggs, N. (2018). Neural machine Translation tools in the language learning classroom: Students' use, perceptions, and analyses. *Jaltcall journal*, *14*(1), 2-24. https://doi.org/10.29140/jaltcall.v14n1.221

Chung, E. S., & Ahn, S. (2021). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer assisted language learning*, *35*(9), 2239-2264. https://doi.org/10.1080/09588221.2020.1871029

Cowie, N., & Sakui, K. (2023). Should machine translation have a role in language classrooms or not? *Pacific journal of technology enhanced learning*, *5*(1), 7-8. https://doi.org/10.24135/pjtel.v5i1.162

Kok Wei, L. (2021). The use of google translate in English language learning: How students view it. *International journal of advanced research in education and society*, *3*(1), 47-53. https://myjms.mohe.gov.my/index.php/ijares/article/view/12459

Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in academic writing courses? *The EuroCALL review*, 26(2), 50–57. https://doi.org/10.4995/eurocall.2018.10140

Lee, S. M., & Briggs, N. (2021). Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL*, *33*(1), 18-33. http://doi.org/10.1017/S0958344020000191

Lee, S.-M. (2020). The impact of using machine translation on EFL students' writing. *Computer assisted language learning*, *33*(3), 157–175. https://doi.org/10.1080/09588221.2018.1553186

Lew, S., Yang, A. H., & Harklau, L. (2018). Qualitative methodology. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 79–101). Palgrave Macmillan.

Tsai, S. C. (2020). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, *35*(5-6), 1250-1272. https://doi.org/10.1080/09588221.2020.1799412

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.17000

Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice

Kevin Papin[a] and Walcir Cardoso[b]
[a]Département de didactique des langues, Université du Québec à Montréal, , papin.kevin@uqam.ca and [b]Department of Education, Concordia University, , walcir.cardoso@concordia.ca

How to cite: Papin, K.; Cardoso, W. (2023). Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.17000

Abstract

This study investigated the pedagogical use of Google Translate's Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) capabilities for improving L2 French students' pronunciation of orthographic {-ent} endings in French: while {-ent} is silent in verbal constructions (e.g. "(ils) résident" [re.zid] '(they) reside'), the same sequence is pronounced [ã] in other forms such as nouns and adjectives (e.g. "(un) résident" [re.zi.dã] '(a) resident'). Twenty beginner learners of French completed self-regulated homework activities using Google Translate to listen to (to develop sound awareness and aural perception) and produce phrases (to develop oral production) containing the target {-ent} forms. A pretest, immediate posttest, and delayed posttest measured awareness, perception, and production of {-ent} pronunciation. Results of repeated-measures ANOVA indicate no significant improvement in awareness or perception, but a significant increase in production scores from pretest to each of the two posttests. The findings suggest that integrating Google Translate's built-in speech technologies into L2 pedagogy can improve the oral production of French grapheme-to-phoneme rules, such as those observed in {-ent} sequences.

Keywords: *L2 French, Google Translate, speech technologies, Automatic Speech Recognition (ASR), Text-To-Speech synthesis (TTS), L2 pronunciation.*

1. Introduction

Although online translation tools such as Google Translate (GT) are becoming increasingly popular among L2 learners for writing practice (Tsai, 2020), CALL researchers have called for an investigation into the potential of their built-in speech technology features to support L2 pronunciation learning (Derwing et al., 2022). These features include Text-To-Speech synthesizers (TTS) and Automatic Speech Recognition (ASR). TTS can provide machine-generated models of speech that learners can imitate and compare against their own pronunciation (Liakin et al., 2015, 2017). ASR, on the other hand, can detect learners' oral production and convert it to text, providing instant feedback on accuracy via the orthographic output (Cardoso, 2018; Garcia et al., 2020). Integrating these speech technologies into L2 pedagogy through platforms like GT could provide personalized, automated feedback to improve the development of pronunciation and speaking skills efficiently (Papin & Cardoso, 2022; van Lieshout & Cardoso, 2022).

Prior research has often examined TTS and ASR separately. These studies found that their pedagogical use can enhance L2 pronunciation (ASR: McCrocklin, 2016; TTS: Liakin et al., 2015). However, only a handful of

studies have targeted L2 French (for exceptions, see Liakin et al., 2017; Papin & Cardoso, 2022) and only three studies have examined the benefits of combining these two technologies in a single L2 learning experience (Khademi & Cardoso, 2022; Papin & Cardoso, 2022; van Lieshout & Cardoso, 2022). Moreover, while Papin and Cardoso (2022) found that the use of GT's ASR and TTS capabilities helped L2 French learners improve their production of French liaison (e.g. the re-syllabification of latent consonants when they appear in consonant-vowel sequences across words, as /z/ in *nos amis* [no.za.mi] 'our friends'), no study has yet looked at the combined impact of using both ASR and TTS on rule-based pronunciation phenomena in French. In this study, we focus on the pronunciation of the orthographic {-ent} sequence, which is pronounced [ã], except in verbal constructions, where it remains silent (e.g. while {-ent} is pronounced [ã] in nouns such as (*un*) *résident* [re.zi.dã] '(a) resident', it is not pronounced in verbal constructions such as (*ils*) *résident* [re.zid] '(they) reside').

The orthographic system of a language can have significant effects on how L2 learners pronounce words in that language (Bassetti et al., 2020; Bürki et al., 2019). Specifically, the regularity and consistency between spelling and pronunciation impacts the ease or difficulty with which learners acquire target pronunciations. The phonology of languages with more regular (transparent) orthographic systems, like Spanish, tend to be easier for learners to acquire than languages with more complex spelling-sound correspondences (opaque), like English. This is because regular orthographies provide clearer cues to guide pronunciation. Irregularities in spelling, on the other hand, can lead learners to mispronounce words (e.g. incorrectly producing [re.zi.dã] instead of the intended [re.zid] in *ils résident*), negatively affecting phonological development. Similar inhibitory effect of orthography on learners' productions has been demonstrated in several studies. For instance, Bassetti et al. (2020) found that Italian learners of English lengthened double consonants (digraphs) of words such as *fi<u>n</u>ish* and *Fi<u>nn</u>ish*, treating each set as minimal pairs in oral production (e.g. producing *fi<u>n</u>ish* with a singleton [n]: ['fɪ.nɪʃ] and *Fi<u>nn</u>ish* with a geminate [n:]: ['fɪ.nːɪʃ]). Given these potential effects, it is important for researchers to closely examine how orthography influences L2 phonological development.

Using a mixed-method approach to data collection, this study addressed the following research question:

- Can the pedagogical use of GT's speech technologies (TTS and ASR) lead to improved pronunciation of the French {-ent} orthographic sequence by French L2 learners?

2. Method

Participants were 20 adult learners enrolled in a beginner L2 French course in an English-speaking Canadian university. To reflect aspects of the participants' pedagogical realities (Erlam & Tolosam, 2022; i.e. the completion of homework assignments), students were asked to complete guided but self-regulated homework activities using GT's TTS and ASR capabilities successively, as part of their course activities. The participants were not presented with any metalinguistic knowledge during the experiment.

The activities targeted the pronunciation assigned to orthographic {-ent} in French, a rule-based orthographic phenomenon that poses difficulty to beginning learners because this morphemic form is homographic in French: while {-ent} is silent in verbal constructions (e.g. (*ils*) *résident* [re.zid]), the same sequence is pronounced [ã] in derived forms such as nouns and adjectives (e.g. (*un*) *résident* [re.zi.dã] '(a) resident').

During the three-week treatment, participants were asked to copy-and-paste a set of phrases containing the target {-ent} forms in GT and press the speaker/playback button (TTS) to listen to the synthesized voice. Participants were then asked to click on the microphone button (ASR) to orally produce the target forms. Finally, they were instructed to verify whether the ASR orthographic output matched what they intended to say (e.g. was intended *(ils) étudient* [e.ty.di] transcribed correctly*?*). The relevant interface of GT and its speech capabilities are illustrated in Figure 1, which also portrays the type of feedback it provides. Here, it displays an incorrect pronunciation for *ils étudient* 'they study', transcribed as produced by the user ([e.ty.djɑ̃] 'student'), not as intended ([e.ty.di]).

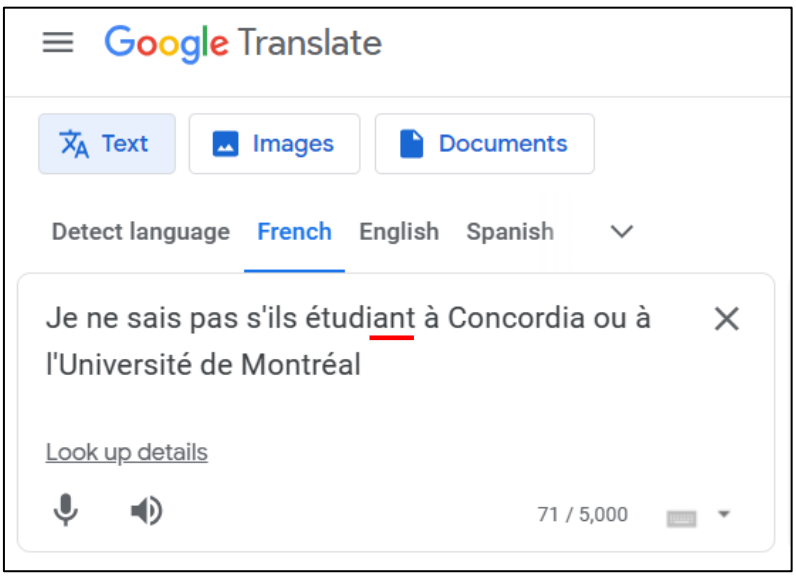


Figure 1. Activity sample in GT: Focus on French {-ent}

To measure the impact of the intervention on the participants' pronunciation development of {-ent}, a pretest and two posttests (immediate and delayed; the latter two weeks after the intervention) were administered. Following Celce-Murcia et al.'s (2010) framework for instructing and evaluating L2 pronunciation, the assessment of phonological development in this study incorporated three distinct levels of analysis (slightly simplified here due to space limitations):

(1) sound awareness (*Awareness*; to assess the participants' level of awareness to the grapheme-to-phoneme rules that characterize homography in {-ent}; e.g. *Is the -ent ending pronounced the same in these two examples: J'ai du talent. vs. Il y a du vent? Can you explain?*);
(2) aural discrimination (*Perception*; to ascertain the participants' ability to differentiate between the two {-ent}forms; participants listened to a short sentence and then decided on whether they heard the target -ent: e.g. *Is the -ent pronounced in the following sentence: La tomate est un aliment bon pour le corps?*); and
(3) oral production (*Production*; to gauge the participants' ability produce the homographic {-ent}; these consisted of reading aloud a set of sentences [e.g. *Les réponses du groupe varient beaucoup*] and a controlled question-and-answer oral task).

3. Results

One-way repeated-measure ANOVAs were performed separately for sound awareness, aural perception, and oral production. The statistical analysis found no significant changes over time for awareness, $F(2, 38) = 2.16$, $p = .287$, partial $\eta 2 = .06$ or perception, $F(2, 38) = .31$, $p = .734$, partial $\eta 2 = .02$. However, the intervention led to statistically significant changes in production over time, $F(2, 38) = 6.52$, $p = .004$, partial $\eta 2 = .26$. Because the study involved multiple comparisons (i.e. three test scores: awareness, perception, production, across three time points: pretest, posttest, delayed posttest), post hoc analysis with a Bonferroni adjustment was conducted, showing a significant increase between the mean scores from pretest to posttest, 1.82 (95% CI, .03 to 3.64), $p = .045$, $d = .50$, and from pretest to delayed posttest, 2.08, (95% CI, .22 to 3.95), $p = .025$, $d = .60$, but no significant increase from posttest to delayed posttest, .27 (95% CI, -.96 to 1.50), $p = 1.00$, $d = .07$. Table 1 summarizes the scores for the three phonological tests at pretest, posttest, and delayed posttest, while Figure 2 provides an illustration of these results (converted to percentages for comparison purposes), emphasizing the significant results in production (where * = $p = .045$ and ** = $p = 025$).

Table 1. Descriptive Statistics: three phonological tests over time (means and standard deviations)

Phonological Levels	Pretest		Posttest		Delayed Posttest	
	M	SD	M	SD	M	SD
Awareness (/3)	1.75	0.72	2.15	0.67	2.00	0.80
Perception (/12)	9.30	1.84	9.55	1.67	9.25	1.67
Production (/25)	15.63	3.14	17.45	4.16	17.71	3.82

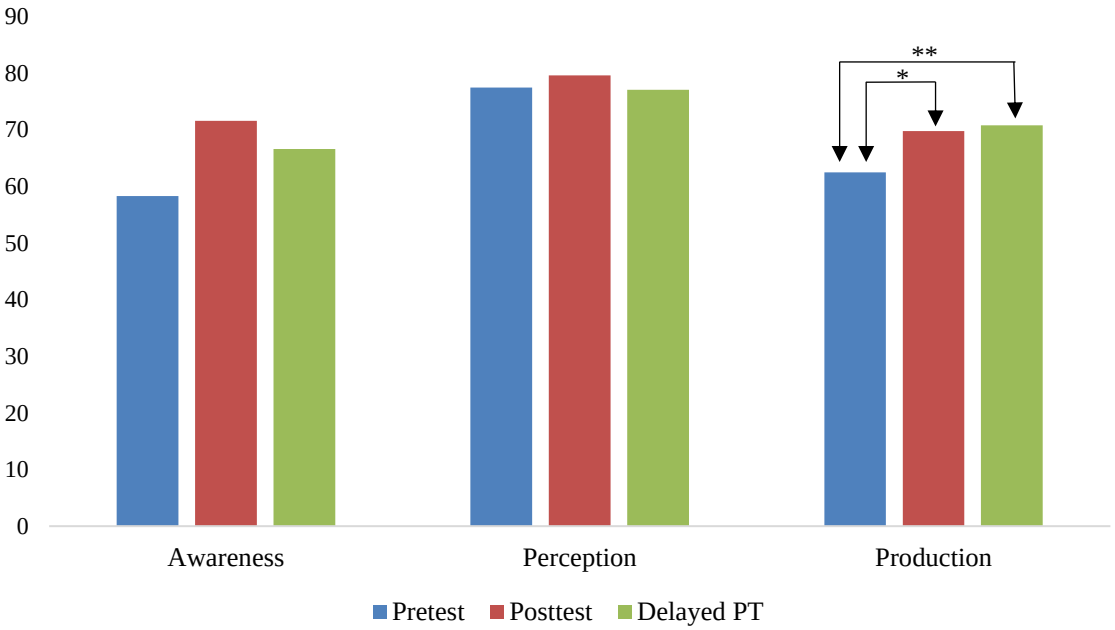


Figure 2. Scores across three tests over time (%).

4. Discussion and Conclusions

This study examined whether the self-regulated use of Google Translate's built-in TTS and ASR could improve L2 French learners' pronunciation of {-ent} endings, an orthographic sequence that is opaque to grapheme-to-phonemes rules: it can be phonetically realized either as [ã] (e.g. in nouns), or remain silent (in verbal forms). As indicated above, the findings only partially confirmed our initial hypothesis: while no gains were detected in the participants development of sound awareness and aural perception, learners' production scores significantly increased from pretest to the two posttests.

These developmental patterns mirror those found for French liaison in Papin and Cardoso (2022), using the same learning platform where the authors found an identical developmental path in which only production benefitted from the proposed treatment, not awareness or perception. In their discussion, the authors attribute this pattern to a ceiling effect in both awareness and perception: since their participants had advanced knowledge of *aural* French liaison before the experiment, indicated on prettest results, there was little room for improvement on those measures compared to production. In the current study, a ceiling effect may also explain the lack of gains

in perception, as participants scored nearly 80% accuracy on the pretest for this measure, not allowing much opportunity for scores to increase further. However, this explanation does not apply for awareness, since pretest awareness scores were comparable to those obtained for production. Specifically, the score for awareness on the pretest was approximately 58.3% (1.75/3), indicating comparatively low knowledge of the {-ent} pronunciation rules. Other factors that may have contributed to the lack of gains in awareness could stem from methodological issues or insufficient practice. For example, the study's measurement of awareness may have been too narrow, failing to capture more subtle gains in explicit knowledge. To assess awareness, one of the questions asked participants to explain the target pronunciation rule. The reliance on metalinguistic explanation as a measure of awareness assumes that this type of explicit knowledge will manifest in verbalizable rules. However, learners may develop awareness that they cannot readily articulate. Additionally, the three-week training period may not have provided the participants enough pronunciation practice of the target {-ent} form so that they could significantly improve in this level of phonological knowledge.

Overall, our findings corroborate those of van Lieshout and Cardoso (2022), whose participants were able to acquire L2 Dutch vocabulary and pronunciation on a short-term basis, also using the two technologies adopted in this study. As such, the results connect to evidence that TTS can provide helpful modeled speech for learning (Liakin et al., 2017), while ASR can offer speaking practice opportunities and instant feedback on learner output (Cardoso, 2018; Liakin et al., 2015). This study uniquely combined these tools for self-regulated practice, operationalized as the completion of homework activities to learn about French pronunciation. As such, it contributes to prior research on utilizing speech technologies for L2 pronunciation instruction.

The study has certain limitations that need to be addressed in future research. The small sample size of 20 learners reduces generalizability. Also, the short-term nature of the pedagogical intervention provides only initial evidence of the adopted tool's capabilities. Additionally, the lack of a control group limits the ability to isolate the pedagogical effects of the target speech technologies. Future research should utilize larger sample sizes, implement a longer-term intervention, and adopt a longitudinal design comparing technology-supported training to control groups undergoing alternative pronunciation instruction approaches.

Acknowledgements

We are grateful to Carol Johnson for her assistance with the statistical analysis. This project received funding from the *Social Sciences and Humanities Research Council* (Papin: 430-2022-00846; Cardoso: 430-2022-00512).

References

Bassetti, B. Mairano, P., Masterson, J., & Cerni, T.. (2020). Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning, 170*(4), 1218-1256. https://doi.org/10.1111/lang.12423

Bürki, A., Welby, P., Clément, M. & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond "friend or foe?". *The Journal of the Acoustical Society of America. 145*(4), EL265–EL271. https://doi.org/10.1121/1.5094923

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds.), *Short papers from EUROCALL* (pp. 165–170). Research-Publishing.net. https://doi.org/10.14705/rpnet.2018.26.806

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). Teaching pronunciation: A course book and reference guide. Cambridge University.

Derwing, T., Munro, M., & Thomson, R. (Eds.). (2022). *The Routledge Handbook of Second Language Acquisition and Speaking*. Routledge. https://doi.org/10.4324/9781003022497

Erlam, R. & Tolosam, C. (2022). *Pedagogical realities of implementing task-based language teaching*. Benjamins. https://doi.org/10.1075/tblt.14

Garcia, C., Nickolai, D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: An empirical study. *Calico Journal, 37*(3), 213-232. https://doi.org/10.1558/cj.40379

Khademi, H., & Cardoso, W. (2022). Learning L2 pronunciation with Google Translate. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022* (pp. 228-233). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1463Papin

Papin, K., & Cardoso, W. (2022). Pronunciation practice in Google Translate: focus on French liaison. *Intelligent CALL, granular systems, and learner data: short papers from EUROCALL 2022* (pp. 322-327). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1478

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

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis: Focus on French liaison. *Computer Assisted Language Learning*, *30*(3-4), 325–342. https://doi.org/10.1080/09588221.2017.1312463

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Tsai, S.-C. (2020). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, *35*(5–6), 1250–1272. https://doi.org/10.1080/09588221.2020.1799412

van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning & Technology, 26*(1), 1–19. http://hdl.handle.net/10125/73460

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16977

An interdisciplinary telecollaboration in ESP classroom: a study involving Business and Law students

W A P Udeshinee[a] and Carmenne Kalyaniwala[b]

[a]Department of Computer and Systems Sciences, Stockholm University, Sweden; Department of English and Modern Languages, NSBM Green University, Sri Lanka, piyumi@dsv.su.se, Piyumi.u@nsbm.ac.lk and [b]University of Lorraine, France, carmenne.kalyaniwala@univ-lorraine.fr

How to cite: Udeshinee, W A P.; Kalyaniwala, C.(2023). An interdisciplinary telecollaboration in ESP classroom: a study involving Business and Law students. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16977

Abstract

Despite the rapid development of telecollaboration in the language classroom, interdisciplinary telecollaborative exchanges in English for Specific Purposes (ESP) classrooms have not received much attention. This paper examines students' perspectives on an interdisciplinary telecollaboration project conducted in the ESP setting. This three-month project involved 23 Law students from France and 24 Business students from Sri Lanka (SL). Their English proficiency ranged from B2 to C1 on the Common European Framework of Reference for Languages scale. They were divided into eight groups. One group consisted of three Business students and three Law students, except the last group, which had only two Law students and three Business students. In each group, Business students had to suggest a business plan to be positioned in Europe, while Law students had to offer legal advice to establish the business in France. They were required to hold at least three Zoom meetings and were allowed to communicate through any other media. Data collected through semi-structured interviews were analysed thematically. The findings revealed that interdisciplinary telecollaboration not only enhances language skills but also soft skills like organisation and time management skills. The paper also discusses the challenges encountered by students and their recommendations.

Keywords: *ESP, Telecollaboration, Online Interactions, Professional Knowledge, Collaborative Learning.*

1. Introduction

Telecollaboration is a project that allows geographically distant parties to engage in authentic communication in a foreign language (O'Dowd & Waire, 2009). These telecollaborative projects have proven to develop learners' linguistic accuracy and fluency (Kinginger & Belz, 2005; Manegre & Udeshinee, 2022), intercultural awareness (Manegre & Udeshinee, 2022), communication skills (García-Sánchez & Gimeno-Sanz, 2022), and digital literacy (García-Sánchez & Gimeno-Sanz, 2022). Such a project promotes the skills that are needed for learners to be competent communicators in the global world. Learners must be equipped with not only language skills but also the social knowledge of how and when to use specific expressions appropriately (Chun, 2015) and the skills required to work in the global market. Therefore, instead of focusing only on second/foreign language learning, it is important to focus on content-based telecollaboration projects (O'Dowd, 2016). English for Specific Purposes (ESP) telecollaboration would provide students with the required practice to work in the professional environment they would be working in future (Lindner, 2015).

Even though there are telecollaboration studies that engage ESP classes, limited research focuses on interdisciplinary virtual exchanges (e.g. Háhn & Radke, 2020; Walker & Brocke, 2009). Further, there are very few telecollaboration projects that involve South Asian universities. Therefore, this study aims to fill this gap by conducting a telecollaboration between Sri Lanka (SL) and France, employing students from two disciplines.

The objective of the project was not only to increase their subject-specific language skills but also soft skills, such as communication and teamwork. Thus, through the analysis of the perceptions of the participants of the project, this study aims to answer the following research questions:

1. Does participating in a telecollaboration improve students' ESP skills?
2. Does participating in a telecollaboration improve students' soft skills?

2. Method

2.1. Context and participants

The study employed 23 Law students from a French university and 24 Business students from a Sri Lankan university. Their age ranged from 18 to 25 years, and English proficiency ranged from B2 to C1 on the Common European Framework of Reference (CEFR) scale.

2.2. Procedure

Business students were expected to suggest a business plan (aligned with Sustainable Development Goals) to be positioned in Europe. Law students were expected to provide legal advice to establish the business in Europe because the SL students were unfamiliar with European legal procedures. For this three-month telecollaboration, eight groups were formed. One group consisted of 3 Business students and three Law students, except the last group, which had only two Law students and three Business students. Since both multimodal and multimedia are necessary for communicative performance (García-Sánchez & Gimeno-Sanz, 2022), several means of communication were allowed. They were required to have Zoom meetings for their discussions, and they needed to record three meetings and upload them to the Google folder shared by the researchers. They could also communicate through WhatsApp or any other media which researchers did not have access to. Further, all their draft documents were shared in the Google folder for researchers to access. More details of the project are given in Table 1.

The telecollaboration project was integrated into the assignments of the Legal English module of the French students; therefore, they received marks for completing tasks. However, the Business students did not receive any marks or a grade. Still, they were given a certificate for participation, and the best business plan and promotional video were selected at a competition held at the SL university.

Table 1. The weekly project plan.

Week	French Team	Sri Lankan (SL) Team
1	Introduction to the project/local teams created	
2	Preparing questions to get to know their partners (cultural information) Recording introductory local team video and uploading it to Soqqle	
3	Workshop on how to think legally: Structures of "social entrepreneurship"	Workshop on Sustainable Development Goals – by an expert

4	Meeting 1	
5	Present information in class collected from meeting 1	Workshop on video making – by an expert
6	Workshop on creating a binding contract	Workshop on business plans – by an expert
7	Meeting 2	
8	Working on the contract	Working on the business plan
9	Meeting 3	
10	Writing a reflective report on the meetings held. Creating a binding contract with insights from Business students	Working on the business plan with insights from the Law students
11	Working on the promotional video	
12	Promotional video	Final competition among the local teams – Presentations on the Business plan (PPT + promotional video) Judged by a panel of judges

2.3. Data collection and analysis

The data were collected from semi-structured, focus-group interviews with students at the end of the project. The collected data were analysed using thematic analysis (Braun & Clarke, 2006). The themes derived from the data have been listed below in Table 2.

Table 2. Main themes emerged from the data.

Main findings	Themes emerged in the analysis of SL students' perceptions	Themes emerged in the analysis of French students' perceptions
The difference in evaluation methods causes conflicts.	Ungraded tasks resulted in less motivation among students.	Challenges encountered due to less interest among some of the SL students
Telecollaboration helped both the English language and soft skills.	The telecollaboration helped improve soft skills more than language skills.	The telecollaboration helped improve speaking and listening skills in English more than ESP skills.
Open-mindedness in participants could overcome the barriers caused by cultural differences.	Cultural differences did not affect the telecollaboration because participants were open-minded.	Cultural differences did not affect the telecollaboration because everyone was open-minded.
Too much work and less explicit instructions could cause challenges.	Telecollaboration amidst other university work and vague instructions posed challenges.	Too much work and insufficient instructions created stressful situations.
Telecollaboration helped gain cultural competence.	Telecollaboration helped learning new things about French culture.	Telecollaboration helped in understanding the social, cultural, and political situation of the country.
The differences in work schedules posed challenges to the effectiveness of telecollaboration.	Different deadlines for the two parties resulted in conflicts.	Different deadlines for the two parties resulted in conflicts.
The different tasks for two parties could reduce the collaboration.	The difference in some tasks reduced the collaboration between the two parties.	The difference in some tasks reduced the collaboration between the two parties.

3. Results and Discussion

3.1. Improvement of ESP skills

The students believed they did not learn any specific business or legal English terms because they were already familiar with the terms used in the discussions. However, they acknowledged that they improved their general English language skills and gained some insights into the business and legal aspects of the project.

"We had to rephrase what we were saying to make them understand what we say, because we didn't have the same level of vocabulary like them, so that improved our English" – A French student

However, as García-Sánchez & Gimeno-Sanz (2022) highlight, learning ESP entails not only learning English specific to the profession but also the improvement of communication skills that will equip them to become international citizens. Some students whose language was already at the C1 level of the CEFR did not necessarily improve their language skills, but their organisation skills, time management skills, business knowledge, legal knowledge, financial knowledge, communication skills (minimising misunderstandings), convincing skills, and video-creation skills.

"Well, rather than English language skills, I learned how to manage time and get the work done without getting stressed. Also, how to work with people with different views " – An SL student

This study also corroborates the findings of Guarda (2013), that engaging non-native speakers of the target language results in less anxiety among students.

"Since English is not their first language, we also felt comfortable in using the language, to make mistakes" – An SL student

In terms of ESP skills, some students noted an improvement in their language skills, while some claimed the improvement in soft skills more than language skills. Thus, it indicates a modest impact of telecollaboration on their ESP skills. However, the reason for this difference could be the varying CEFR levels among students. If the study used elementary-level students, the results would have been different, but more research is needed to prove this claim.

3.2. Improvement of cultural competence

Students of both cultures revealed that the culture posed no barriers when communicating with each other as all the members were open-minded. In line with many previous studies (e.g. Lee, 2019; Manegre & Udeshinee, 2022), this study revealed that telecollaboration improved their cultural competence, which supports them to become global citizens.

"Collaboration is all about understanding each other and their expectations and coming to a middle point"- SL student.

SL students mentioned their perception of French culture changed after the collaboration because they had previously visualised its culture as it was portrayed in popular movies and TV series. Although SL students knew about the French culture more than French students did about Sri Lankan culture, they still learned new things about it. Similarly, French students obtained some insights into Sri Lankan people and culture through collaboration.

3.3. Challenges of telecollaboration

The project also encountered challenges. Confirming the findings of Helm (2015) and O'Dowd & Ritter (2006), this study reveals that lack of student motivation, the difference in aims and approaches to the exchange, students' workload, and different timelines were challenges encountered in the project. There seemed to be some conflicts between the two groups, mainly because they did not share the same purpose, leading to less motivation in some SL students.

"We were not stressed about the presentations, because we were not given marks, we wanted to have fun. Also, you can't experiment with things in assignments, this was a good opportunity to do some experiments, to try out new things" – An SL student

Another challenge highlighted were tight deadlines. Further, different deadlines for each party impeded effective collaboration. The allocation of different tasks, such as Law students handling contracts and Business students focusing on business plans, limited opportunities for collaboration.

3.4. Recommendations for better telecollaboration

Students suggested that the time a student can commit to the project should be considered when recruiting them. They also proposed enhancing flexibility in schedules and providing clear explanations of roles for both parties to enhance the success of the project. Further, SL students suggested that allowing both parties to present at the competition would increase collaboration.

4. Conclusions

This study examined how ESP and soft skills could be improved in interdisciplinary telecollaboration. The study's findings indicated that telecollaboration supported the students who were at the C1 level especially with improving their soft skills, while helping the rest of the students to improve both language and soft skills. The study had some limitations. First, the evaluation method used for the two parties was not the same. Sri Lankans were given a certificate, and the French were given marks; this may have affected the students' motivation level. Second, the deadlines differed for the two parties; this would also have affected the students' motivation and engagement. However, it would also be interesting to examine the results of such a telecollaboration, which includes the same evaluation method and deadlines. Finally, a greater collaboration could have been fostered if both parties were involved in the same evaluation process.

Acknowledgements

We would like to thank all the students who participated in the study and Loic Palmisano of the University of Loraine for his support in implementing the project. Further, we are grateful to all the experts who shared their knowledge with the students.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Chun, D. M. (2015). Language and culture learning in higher education via telecollaboration. *Pedagogies: An International Journal*, *10*(1), 5–21. https://doi.org/10.1080/1554480X.2014.999775

García-Sánchez, S., & Gimeno-Sanz, A. (2022). Telecollaborative Debates in Esp: Learner Perceptions and Pedagogical Implications. *Journal of Teaching English for Specific and Academic Purposes*, *10*(1), 139–157. https://doi.org/10.22190/JTESAP2201139G

Guarda, M. (2013). *Negotiating a transcultural place in an English as a lingua franca telecollaboration exchange*. http://paduaresearch.cab.unipd.it/5337/1/guarda_marta_tesi.pdf

Háhn, J., & Radke, K. (2020). Combining expertise from linguistics and tourism: a tale of two cities. In F. Helm & A. Beaven (Eds.), *Designing and implementing virtual exchange - a collection of case studies* (Issue 2020, pp. 11–22). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.45.1111

Helm, F. (2015). *THE PRACTICES AND CHALLENGES OF TELECOLLABORATION IN*. *19*(2), 197–217.

Kinginger, C., & Belz, J. A. (2005). Socio-cultural perspectives on pragmatic development in foreign language learning: Microgenetic case studies from telecollaboration and residence abroad. *Intercultural Pragmatics*, *2*(4), 369–421. https://doi.org/10.1515/iprg.2005.2.4.369

Lee, J. (2019). *Developing Intercultural Competence through Study Abroad , Telecollaboration , and On-campus Language Study*. *23*(3), 178–198.

Lindner, R. (2015). Developing communicative competence in global virtual teams : A multiliteracies approach to telecollaboration for students of business and economics. *CASALC Review*, *5*(1), 144–156.

Manegre, M., & Udeshinee, P. (2022). A telecollaborative study of university students in Spain and Sri Lanka using the Soqqle video app. *Intelligent CALL, Granular Systems, and Learner Data: Short Papers from EUROCALL 2022, 2022*, 252–257. https://doi.org/10.14705/rpnet.2022.61.1467

O'Dowd, R. (2016). Learning from the Past and Looking to the Future of Online Intercultural Exchange. In R. O'Dowd & T. Lewis (Eds.), *Online Intercultural Exchange: policy, Pedagogy, Practice*. Routledge.

O'Dowd, R., & Ritter, M. (2006). in Telecollaborative Exchanges. *CALICO Journal*, *23*(3), 623–643.

O'Dowd, R., & Waire, P. (2009). Critical issues in telecollaborative task design. *Computer Assisted Language Learning*, *22*(2), 173–188. https://doi.org/10.1080/09588220902778369

Walker, U., & Brocke, C. (2009). Integrating content-based language learning and intercultural learning online : An international eGroups collaboration. In A. Brown (Ed.), *Proceedings of CLESOL 2008*.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16932

Cultural identity construction and English as a lingua franca in online virtual exchange

Alice Gruber[a] and Sofiya Karnovska[b]

[a]Center for Languages and Intercultural Communication, Augsburg Technical University of Applied Sciences, , alice.gruber@hs-augsburg.de and [b]Center for Languages and Intercultural Communication, Augsburg Technical University of Applied Sciences, , sofiya.karnovska@hs-augsburg.de

How to cite: Gruber, A.; Karnovska, S. (2023). Cultural identity construction and English as a lingua franca in online virtual exchange. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16932

Abstract

This study explores how students perceive the construction of cultural identity in online intercultural encounters and the impact these perceptions may have on their relationship with English as a Lingua Franca (ELF). This paper presents findings from a mixed-methods study conducted with university students in Germany, France, Israel, and Spain, focusing on English language learning experiences within the context of a Virtual Exchange (VE) project. Results from qualitative data of (n=356) participants indicate considerable differences in how students perceive their roles in intercultural interactions, along with the benefits associated with ELF participation. The results moreover emphasise the importance of reflective practices and open discussions concerning identity construction and language use. The value of initiating dialogues and reflecting identity in the ELF classroom is discussed.

Keywords: *English as a Lingua Franca (ELF), identity construction, intercultural communication, Virtual Exchange (VE).*

1. Introduction

Cultural identity as a subdomain of social identity has yet to be explored in VE settings. A common assumption is that speakers in lingua franca conversations are "representatives of their individual mother cultures" (Meierkord, 2000, n.p.). However, it is not clear to what extent participants in VE projects perceive themselves as cultural ambassadors rather than as individuals. To answer the question *Who am I?*, individuals construct complex, multi-faceted, and shifting identities consisting of personal and social identity domains (Vignoles et al., 2011). The social identity domain is a broad concept encompassing cultural, national, regional, and other group affiliations and plays an important role in communication and social interaction (Byram, 2023). Social factors, such as an individual's affiliations with a wide range of social groups based on, for instance, gender, social class, ethnicity, and language (e.g., Deaux, 2018; Kramsch, 2013), are a central aspect of identity construction. Therefore, the effect of VE projects on participants' cultural identities poses a research interest.

Although people frequently regard themselves and others as distinct individuals, there are many situations in which they think, feel, and act primarily as members of a social group, and where collective concerns take precedence over individual differences (Ellemers, 2012). Social contexts that require collaboration, such as those created by VE projects, offer students the opportunity to construct, negotiate, explore, and reflect on their identity when collaborating with people from different linguacultural backgrounds.

VE projects that use ELF as the means of transcultural communication provide authentic communication opportunities on an equal footing (Hoffstaedter & Kohn, 2015) or a potentially more equal footing than in a Native-Speaker (NS)-Non-Native Speaker (NNS) constellation. Generally, learners in ELF communication appear to feel more confident and less concerned about making errors, as the emphasis is on communicating effectively (Nicolaou & Sevilla-Pavón, 2023). This paper reports on a mixed-methods study conducted with ELF students as part of a VE project and examines student perceptions of cultural identity construction and ELF.

2. Method

The study was conducted during a six-week mentored VE project in the spring of 2023. A total of 356 university students from Chile, Israel, France, Spain, and Germany participated in the VE, which was offered as part of students' regular English classes. The participants were enrolled in Bachelor's, Master's, and doctoral programs at their respective universities, studying a range of courses (e.g. Business Studies and Software Engineering). The students' English proficiency ranged between A2 and C2. The groups were composed of three or four students on average. At the beginning of the project, the participants gave their consent to participate in the study. The participants then responded to an online survey containing demographic questions and a suite of validated questionnaires, e.g. the Cultural Intelligence Scale (Ang et al., 2007). After the VE, student perception of the VE experience was gauged using a post-project questionnaire that included open-ended questions about their cultural identity and relationship to English and ELF communities in general.

In total, there were 356 post-project questionnaire responses. In this study, we report on the qualitative analysis of open-ended questions from the post-intervention questionnaire based on a thematic analysis approach (Braun & Clarke, 2006). The aim is to explore the following Research Questions (RQs):

1) To what extent do students who participate in a VE project using ELF feel they represent themselves as individuals rather than representatives of one or more cultures?
2) What were the students' perceptions of interacting with other non-native English speakers during a VE project?
3) In what way did the VE influence students' relationship with the English language?

3. Results

3.1. Results concerning RQ1

151 participants responded to RQ1, *Did you perceive yourself as an individual rather than a representative of one or more cultures?* Seven students who completed the questionnaire chose not to answer this question, which might indicate that they felt uncomfortable or found the question difficult. The distribution of participants' answers can be seen in Figure 1.

42% (n=63) of participants confirmed that they perceived themselves solely as individuals during the exchange. Several highlighted the uniqueness and diversity of individuals as the primary reason for their inability to represent the culture(s). The following statements provide insights into the students' reasoning behind their perspective:

- "Each person is unique, and I cannot be responsible for the whole culture" [P232].
- "I'm just an individual, so I cannot speak for a whole country" [P190].
- "Why should I represent 83 million others?" [P34].

Within a student group composed of at least two individuals from the same cultural background (e.g. two out of four students from Germany), an Israeli student noted the significance of individual differences between himself and another Israeli student:

- "I feel that I represented myself and my opinions only. Of course, I did talk about my country and religion, but mostly I emphasized my opinions. You can see the differences in thought between me and another Israeli participant; we are both Israeli students, and each has a different opinion and represents something different" [P145].

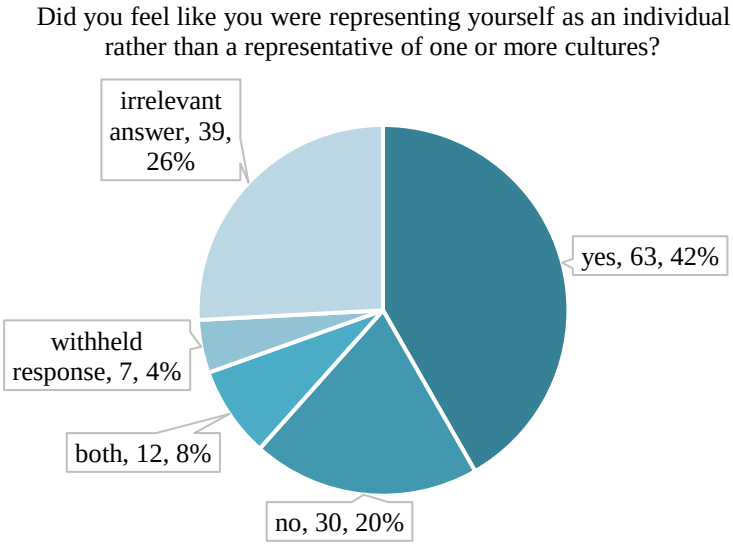


Figure 1. Distribution of answer codes to RQ1 (n=151).

Some students wanted to clearly indicate their differences or distance themselves from their culture(s):

- "I did explain why I couldn't identify with certain stereotypes or where the difference is between the common behavior in my culture and my behavior" [P219].
- "I represented my values more than the values of our whole culture, as I see some things differently than other members of our culture" [P253].

20% (n=30) of students expressed a strong sense of cultural representation. Some students felt responsible for sharing valuable insights and knowledge about their culture(s):

- "[The exchange partners] only know us as representatives of our culture. And they aren't able to know which of our behaviour is culture-related and which is individual" [P37].
- "They do not know a lot about my culture, so I have to explain it in the best possible way" [P219].
- "Me as a representative of my culture because my traditions and beliefs belong to my culture" [P358].
- "One student assumed that everyone in their group was representing their culture"
- "Everyone in my group understood himself as a representative; I hopped into my role" [P275].

A couple of multicultural students explicitly commented on representing or not being able to represent more than one culture:

- "I felt like I represented several cultures" [P171].
- "[I did not feel like a representative] because I don't know everything about my two cultures" [P50].

8% (n=12) felt they represented both themselves and their culture. One student reasoned: "Both, because I am part of German culture and I grew up here, but of course, I am an individual person" [P233]. Another student felt it was contextual: "Both and depends on the situation" [P329]. P257 reflected: "In the meetings, I mixed my personal experiences and preferences with the general behaviour of German people. That is why sometimes I felt both like an individual and a representative."

To see how cultural background may relate to feeling like a representative of an entire culture, the distribution of responses to the questions PP07 ("Did you feel like a representative of an entire culture when talking to your partner(s)?") and BG01 ("Do you have a pluricultural background?") were checked for interaction with a Pearson's Chi-Square test in SPSS (version 29). The Chi-Square test results (X^2 (2,150)=1.71, p = .43) indicated that BG01 and PP07 were statistically independent from one another. Consequently, cultural identity constellation (BG01) and feeling of cultural respresentation (PP07) did not interact.

3.2. Results concerning RQ2

153 participants responded to RQ2, *How did interacting with other non-native speakers of English change your view of people who don't speak English as their first language?* The distribution of answers is illustrated in Figure 2.

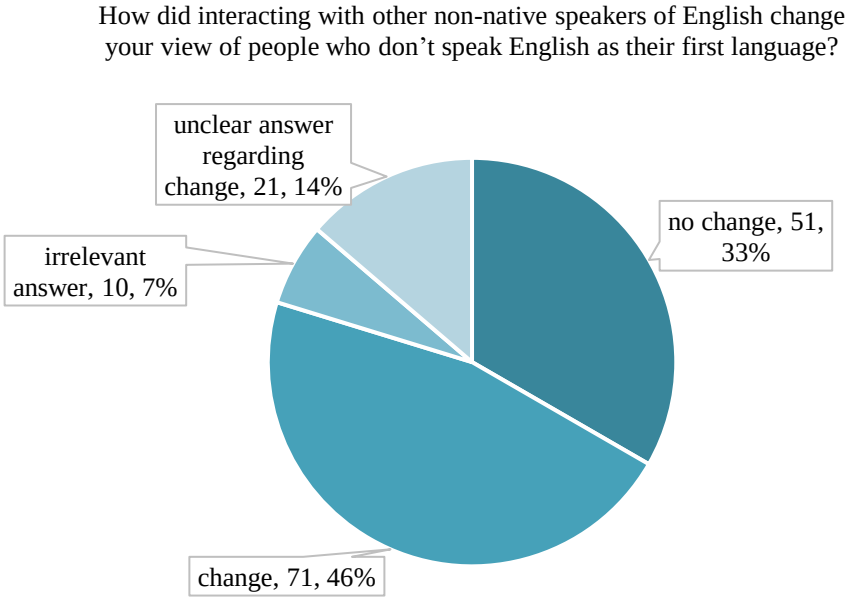


Figure 2. Distribution of answer codes to RQ2 (n=153).

Out of 153 participants, 33% (n=51) explicitly stated that there had been no change in their views on NNS. 9% (n=14) stated that they were already used to speaking to other NNS as a reason for not changing their views.

The data showed that students compared their proficiency among the whole group (e.g. "Level of their English was quite the same as ours" [P264]; "We Germans do speak better English than our counterparts from Spain" [P34]) or individually (e.g. "It's good to know that it's not only me having a problem with speaking English perfect fluently" [P191]; "They are not better in speaking than me" [P 289]).

46% of students' responses to the question indicated a change in attitude, involving more awareness and acceptance of NNS. This is reflected in the following statements:

- "Understanding that they might have an accent or make grammar related mistakes (just being respectful and trying to help if someone misses a word)" [P219].
- "Pardon small mistakes" [P360].
- "It doesn't matter as long as you are kind" [P37].
- "It's ok to make mistakes" [P209].

One student reflected on the advantage of a multilingual approach: "We all got the same problems by learning the language. It's easier to communicate if we mix the languages" [P49]. In the context of the ELF constellation, students mentioned increased motivation to improve their language skills ("Gave me the drive to improve my English so I won't be misunderstood" [P344]) and lower anxiety (e.g. "You do not feel like you are under pressure" [P260]; "I felt more safe because I saw that people from other countries do not speak English perfectly" [P202]).

3.3. Results concerning RQ3

135 participants responded to RQ3, *In what way did the VE affect your relationship with English?*. 91 students (59%) described a more positive attitude towards English because of the VE, whereas 27% reported no change in their relationship with English. Only one student reported a more negative attitude towards English because of the VE. Figure 3 shows the distribution of the responses.

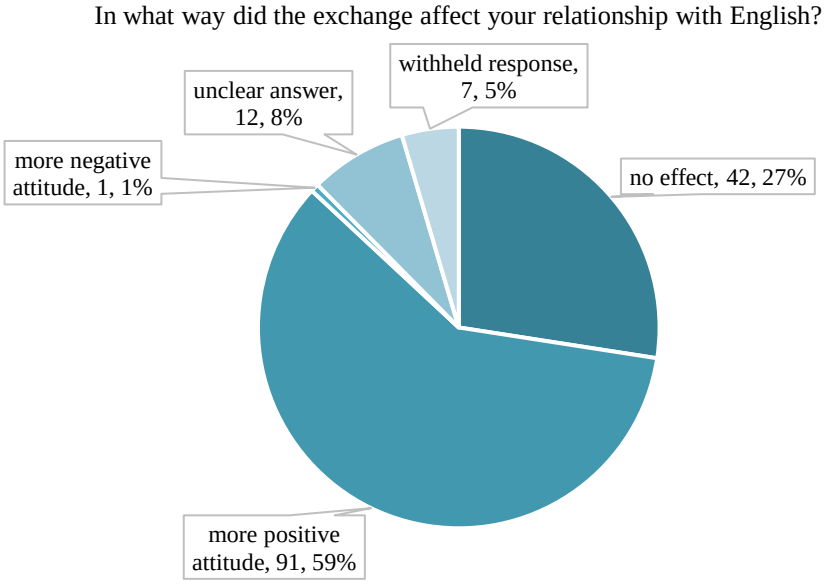


Figure 3. Distribution of answer codes to RQ3 (n=153).

54 students (35 %) explicitly commented that they felt more confident, comfortable, fluent, or safer because of the VE. Other students commented that the VE was useful in helping them realise that they need to enhance their English skills.

- "I realised that I need to work on my communication skills and vocabulary in discussions" [P186].
- "I recognized that English is important to communicate with other people I want to improve because it was sometimes not fluent" [P45].

4. Discussion and conclusions

This study explored students' perceptions of ELF in a VE project and its impact on their (cultural) identity construction. The research findings for RQ1 demonstrated the complexity and diversity of the students' cultural identities and identity attitudes in the VE setting. A possible reason for this range of perspectives could be the students' cultural background (e.g. individualist vs collectivistic culture). This and other possible reasons should be explored in future VE research projects.

In answering RQ2, some students mentioned not speaking 'perfect English', implying an ideologically monolithic approach to English. Nonetheless, the answers also show that the VE project improved some students' levels of

tolerance towards other NNS. Additionally, the project and the questions encouraged self-reflection on language proficiency and English perspectives.

The analysis for RQ3 showed that many students felt more confident, comfortable, fluent, or safer and were more motivated to improve their English because of the VE. This aligns with previous research results (e.g. O'Dowd, 2021). A more positive relationship with English suggests VEs can be effective in enhancing learner attitudes and motivation. The findings show that in VE projects involving ELF-speaking participants, students' perceptions of their role and the benefits they derive from the ELF constellation vary considerably. In terms of pedagogical implications, engaging students in reflective discussions and reflections of ELF, including myths such as 'perfect English', language varieties, and traditional assumptions perpetuating Standard Language within guided VE projects is valuable. Materials to prepare themselves for such discussions are available to teachers (Hall et al., 2022). Similarly, prompting discussions about students' cultural identities can be beneficial to potentially gain a deeper understanding of their identities. The types of guidance and materials for university students and teachers needed in VEs require more exploration. Further research should examine the extent to which students' self-reported views align with their behaviours during VEs. Moreover, the potential impact of the medium (e.g. a videoconference, the physical classroom, or virtual reality environments) on students' identity construction should be investigated.

References

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. *Management and organization review*, *3*(3), 335–371. https://doi.org/10.1111/j.1740-8784.2007.00082.x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Byram, M. (2023). Identity matters in the Common European Framework of Reference for Languages and its Companion Volume. *The Language Learning Journal*, *51*(2), 253–262. https://doi.org/10.1080/09571736.2021.1996622

Deaux, K. (2018). Ethnic/Racial Identity: Fuzzy Categories and Shifting Positions. *The ANNALS of the American Academy of Political and Social Science*, *677*(1), 39–47. https://doi.org/10.1177/0002716218754834

Ellemers, N. (2012). The group self, *Science* 336(6083), 848–852. https://doi.org/10.1126/science.1220987

Hall, C., Gruber, A. & Qian, J. (2022). Modelling plurilithic orientations to English with pre-service teachers: A comparative international study. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3181

Hoffstaedter, P., & Kohn, K. (2015). Telecollaboration for intercultural foreign language conversations in secondary school contexts: Task design and pedagogic implementation. *TILA Research Results on Telecollaboration*.

Kramsch, C. (2013). Culture in foreign language teaching. *Iranian Journal of Language Teaching Research*, *1*(1), 57–78.

Meierkord, C. (2000). Interpreting successful lingua-franca interaction. An analysis of non-native-/non-native small talk conversation in English. *Linguistik Online*, *5*(1). https://doi.org/10.13092/lo.5.1013

Nicolaou, A., & Sevilla-Pavón, A. (2023). Exploring the impact of virtual exchange in Virtual English as a lingua franca (VELF): Views on self-efficacy and motivation. In I. Pineda & R. Bosso (Eds.), *Virtual English as a Lingua Franca* (pp. 173–194). Routledge.

O'Dowd, R. (2021). What do students learn in virtual exchange? A qualitative content analysis of learning outcomes across multiple exchanges. *International Journal of Educational Research*, *109*, 1–13. https://doi.org/10.1016/j.ijer.2021.101804

Vignoles, V. L., Schwartz, S. J., & Luyckx, K. (2011). Introduction: Toward an Integrative View of Identity. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of Identity Theory and Research* (pp. 1–27). Springer New York. https://doi.org/10.1007/978-1-4419-7988-9_1

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16925

Enhancing university EFL students' informal reasoning on the social scientific issues related to sustainable development goals by adopting a collaborative argumentation CALL environment

Li-Jen Wang[a]

[a]Language Teaching and Research Center, National Yang Ming Chiao Tung University, , tommywang@nycu.edu.tw

How to cite: Wang, L.-J. (2023). Enhancing university EFL students' informal reasoning on the social scientific issues related to sustainable development goals by adopting a collaborative argumentation CALL environment. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16925

Abstract

In the 21st century, educators need to develop students' argumentation skills for addressing Social Scientific Issues (SSIs). The Sustainable Development Goals (SDGs) proposed by the United Nations (UN) align well with SSIs argumentation. Informal reasoning plays a crucial role in this process, but existing research focuses mainly on individual learners and subjective opinions. Recently, the collaborative learning environment of Knowledge Forum has been highlighted as a potential enhancer of students' informal reasoning, but its effectiveness in language teaching remains understudied. Thus, this study examined a collaborative argumentation Computer Assisted Language Learning (CALL) environment to enhance informal reasoning skills of university English as a Foreign Language (EFL) students. The experimental group (n=25) utilized the collaborative argumentation CALL environment with Knowledge Forum, while the control group (n=30) used a conventional argumentation environment. The results showed that the experimental group outperformed the control group in informal reasoning quality, including counterarguments and rebuttals. These findings suggest that collaborative argumentation CALL environments may enhance EFL students' informal reasoning, providing valuable insights for educators seeking to improve students' informal reasoning skills.

Keywords: *CALL, collaborative argumentation, informal reasoning, social scientific issues, sustainable development goals.*

1. Introduction

In the 21st century, a multitude of challenges has arisen as a result of human activities. To address these challenges, the UN introduced the SDGs in 2015. These challenges, commonly referred to as SSIs (Azumah & Marlizayati, 2023; Nilay & Ozgul, 2017; Sadler, 2004; Sadler & Zeidler, 2005; Wu & Tsai, 2007), are characterized by their reliance on scientific concepts or problems, controversial nature, public discourse, and the influence of political and social factors (Sadler & Zeidler, 2005, p.113). Within educational settings, researchers have emphasized the significance of developing university students' informal reasoning skills to facilitate their ability to discuss and address SSIs (Sadler, 2004; Sadler & Zeidler, 2005; Wu, 2013). Furthermore, university students are expected to engage in negotiation and resolution of contentious issues, such as SSIs, by providing supporting evidence for their claims and refuting counterarguments. Consequently, informal reasoning serves as a valuable tool for discussing SSIs within university classroom settings (Cerbin, 1988; Kuhn, 1993). From the perspective of SDG-related SSIs, the majority of empirical studies on SSIs argumentation have primarily focused on examining

relevant issues from the standpoint of individual learners. However, only a limited number of studies have investigated SSIs argumentation from the angles of social interaction, dialogue theory, or cooperative learning (Azimah & Marlizayati, 2023). Training in collaborative argumentation highlights the potential of adopting the knowledge-building theory as an effective teaching approach. This approach encourages collaboration, dialogue, and a puzzle-like learning approach that enables university students to share and co-create knowledge from various perspectives. To enhance the comprehensibility and applicability of knowledge-building in education, Scardamalia (2002) proposed twelve principles of the knowledge-building theory and her team developed an online collaborative knowledge-building platform called *Knowledge Forum*.

The Knowledge Forum platform aims to establish a collaborative environment for knowledge building, enabling university students to contribute with their ideas, share them, and engage in discussions with fellow students. What sets this platform apart from general online discussion forums or web blogs is its incorporation of built-in knowledge-building scaffolds. These scaffolds assist users in evaluating the attributes of their speech and selecting appropriate scaffolds as annotations before posting or responding. To put it briefly, this knowledge-building, theory-based platform can function as an online collaborative argumentation online platform, aiding different kinds of students in improving their informal reasoning abilities in various subjects. So far, there has been no research conducted in the domain of language learning regarding this matter. In the context of higher education, the development of EFL university students' ability to propose potential solutions to SDG-related SSIs is of paramount importance. Nevertheless, the integration of SSI-based argumentation in EFL classrooms has not received substantial research attention nor has it been adequately incorporated into EFL university courses. The objective of this study is therefore to investigate whether EFL university students can utilize Knowledge Forum as a collaborative argumentation CALL environment for engaging in informal reasoning on SDG-related SSIs. The research question in this study is: "Do university students perform better informal reasoning skills in the collaborative argumentation CALL environment?"

2. Method

2.1. Research design and participants

This study employed a quasi-experimental research methodology. The participants consisted of two classes (n=55) enrolled in an 'English Reading and Writing' course at a university in Taiwan in 2022. The duration of the course was 16 weeks, with two hours per week, and it was taught by the same experienced instructor who has been incorporating the CALL environment into the curriculum. All participants, including 35 males and 20 females, were Mandarin native speakers (L1) with similar levels of English proficiency (CEF B2 to C1). They were at least 20 years old and provided written consent. They were randomly assigned to either the experimental group (n=25) or the control group (n=30). The research design is presented in Figure 1. The study spanned an eight-week period, representing half of the course duration. Both the experimental and control groups underwent a pre-test in week one and a post-test in week eight to assess their Informal Reasoning (IR) skills before and after the intervention. Both the pre-test and the post-test implemented the same open-ened informal reasoning performance paper questionnaire, which was sent to the participants in class to investigate the changes of the participants' informal reasoning skills (see section 2.3.). In weeks 2 and 3, the experimental group received instruction on collaborative argumentation, focusing on knowledge-building theory and operational training utilizing the Knowledge Forum platform. In contrast, the control group received conventional argumentation instruction through teacher-led lectures. During this phase, both groups were assigned an SDG-related SSI topic (specifically, the nuclear power issue, linked to SDG 7: affordable and clean energy). Additionally, all participants were assigned various learning tasks, such as reading articles on the nuclear power issue and learning how to retrieve credible online information. Weeks 4 to 7 encompassed classroom-based discussions on the assigned topic, with the two groups operating in different environments. The experimental group engaged in collaborative argumentation within the CALL environment, utilizing the Knowledge Forum platform for their collaborative discussions. On the other hand, the control group participated in conventional argumentation discussions.

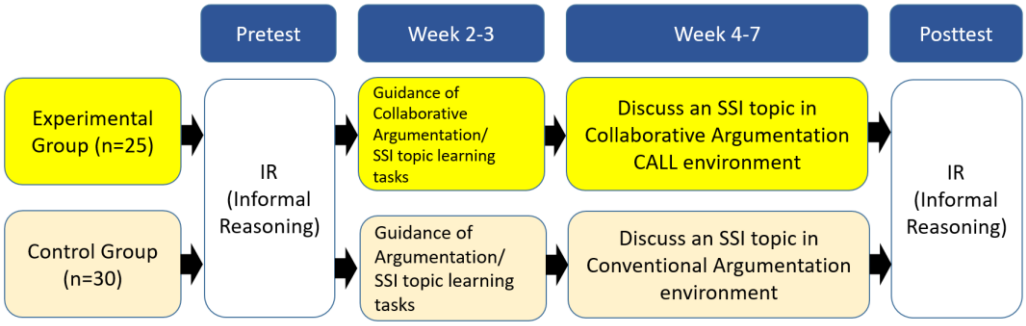


Figure 1. Research design.

2.2. Two learning environments for discussing the SDG-realted SSI

In this study, the collaborative argumentation CALL environment was facilitated by the Knowledge Forum platform, providing a multimedia-based knowledge space for community members to contribute ideas and enhance their initial understandings (Hong et al., 2014; Hong & Chiu, 2015; Sun, Zhang, & Scardamalia, 2010). It is worth noting that the scaffolds aid students in clarifying and organizing their conceptual writings in notes. Figure 3 demonstrates the six predefined scaffolds integrated into the platform. Students were required to employ the scripted scaffolds embedded in the platform to explain their purposes for responding (see Figure 2). This process was similar to informal reasoning. In contrast, the control group engaged in face-to-face group discussions on the SSI topic within the conventional argumentation environment, without utilizing any online platforms.

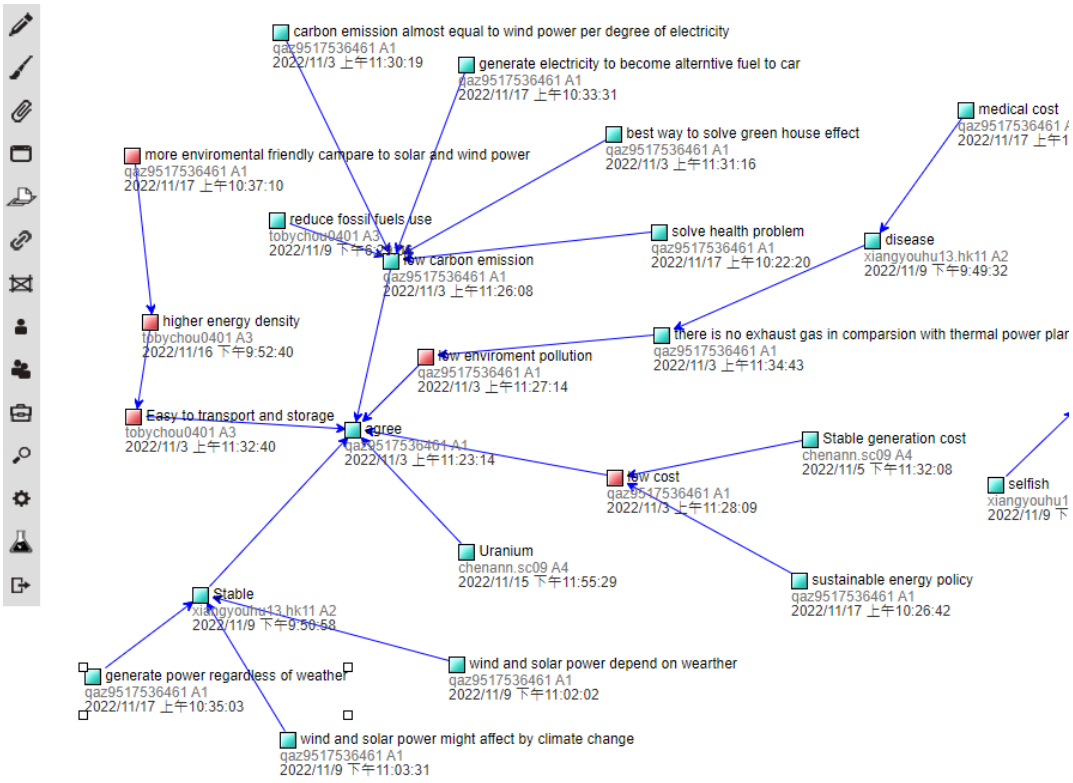


Figure 2. Students' Knowledge Forum notes.

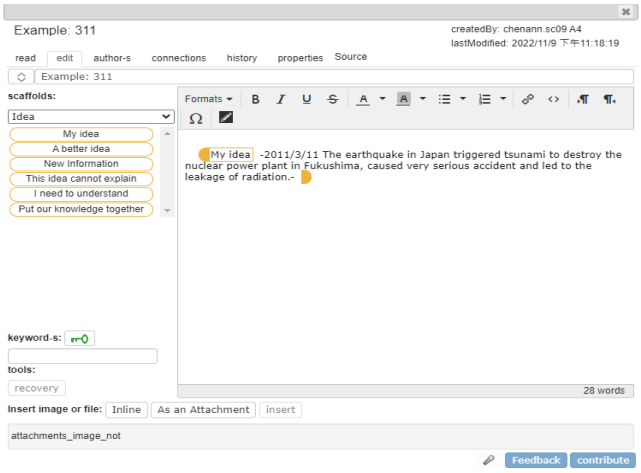


Figure 3. Knowledge Forum embedded scaffolds.

2.3. Instruments

An open-ended informal reasoning performance questionnaire developed by Wu and Tsai (2013) incorporating elements of argumentation was adapted to collect data and to evaluate students' informal reasoning on the SDG-related SSI topic. The questionnaire consisted of four questions, each serving a different purpose: Q1:*Understanding claim: "Do you agree or disagree with the use of nuclear power to address Taiwan's power supply problem?"*; *Q2: Proposing arguments: "Please provide your arguments and evidence supporting your position on this issue."*; *Q3: Considering counterarguments: "If someone holds a different position from yours, what counterarguments might be raised to challenge your arguments in the previous question?"*; *Q4: Rebuttals: "How would you rebut the counterarguments with supporting evidence?"*

2.4. Data collection and analysis

The pre-test and post-test responses were analyzed qualitatively and quantitatively using the analytical framework adapted from Wu and Tsai (2007). They developed an informal reasoning quality indicator for their analytical framework, as shown in Table 1. Further descriptions are provided in Table 2. The data from participant students' responses were read and coded separately by the author and an experienced colleague. An initial agreement between the raters was 83%. A 100% agreement rate was eventually achieved between the two raters after several face-to-face and virtual discussions to clarify their understanding of the categories.

Table 1. Description of reasoning quality level.

Level	Description
None	Only claim provided
Lower	Claim and supportive argument OR counterargument provided
Medium	Claim, supportive argument, and counterargument provided
Higher	Claim, supportive argument, counterargument, and rebuttal provided

Table 2. Description of claim, supportive argument, counterargument, rebuttal, and evidence.

Structure	Description
Claim	To show stance, proposition, or assertion about an issue
Supportive argument	The statement contains reasons and evidence to support a claim
Counterargument	Alternative assertion to a person's claim with accompanying evidence
Rebuttal	Valid rejection of a reason that is in support of a counterargument with accompanying evidence

Evidence	Evidence can be from students' knowledge or own experience, numerical or descriptive data, and concrete examples or facts

Adapted from Azimah & Marlizayati (2023)

3. Results

As shown in Table 3, after the conduct of this study, both the control group and the experimental group with different argumentation learning environments have shown an increase from the pre-test to the post-test. Table 4 further shows the frequency of informal reasoning quality levels generated by participant students in the pre-test and post-test are presented. It is observed that the none level of the control and the experimental groups decreased slightly from 10 % (pre-test) to 6 % (post-test), and from 8 % (pre-test) to 4 % (post-test) respectively. The lower level of the control group decreased from 40 % (pre-test) to 30 % (post-test), while the experimental group dramatically decreased from 44 % (pre-test) to 8 % (post-test). Moreover, the frequencies of medium level of both the control and the experimental groups increased from 20 % (pre-test) to 33 % (post-test), and from 28 % (pre-test) to 36 % (post-test) respectively. The most remarkable finding is the significant increase of the experimental group in the frequency of higher reasoning quality levels from 20% (pre-test) to 52% (post-test), although the control group has shown the same trend from 20% (pre-test) to 30% (post-test).

Table 3. The frequency of claim and type of argument.

Claim and type of argument	Control group		Experimental group	
	Frequency		Frequency	
	Pre-test	Post-test	Pre-test	Post-test
Claim Agree	26	26	22	22
Claim Disagree	4	4	3	3
Supportive argument	28	32	27	29
Counterargument	14	19	13	25
Rebuttal	6	10	7	25
Total number of argument	48	61	47	79

Table 4. Participant students' reasoning quality level during the pre-test and the post-test.

Reasoning quality level	Control group		Experimental group	
	Frequency (%)		Frequency (%)	
	Pre-test	Post-test	Pre-test	Post-test
None	3 (10%)	2 (6%)	2 (8%)	1 (4%)
Lower	12(40%)	9 (30%)	11(44%)	2 (8%)
Medium	9 (30%)	10 (33%)	7 (28%)	9 (36%)
Higher	6 (20%)	9 (30%)	5 (20%)	13(52%)

4. Discussion

It is found that no matter which argumentation learning environment, our university students did improve their frequency of claim and type of argument. It is noted that students performed better informal reasoning skills in the collaborative argumentation CALL environment in terms of the frequency of claim and type of argument, and their reasoning quality. The experimental group could generate more arguments. This finding confirms that the collaborative argumentation CALL environment did advance students' knowledge and discourse and assist them in generating and refining their ideas (Hong & Scardamalia, 2015). In terms of the level of reasoning quality, a noticeable decline in the none level and lower level of reasoning quality, as well as an increase in the higher level of reasoning quality, can be observed in both the control and experimental groups. Scholars have posited that constructing counterarguments and rebuttals poses a cognitive challenge for university students (Erduran, et al.,

2004; Azimah & Marlizayati, 2023). It is noteworthy that students in the experimental group exhibited a tendency towards achieving higher reasoning quality compared to the control group. This observation potentially validates the effectiveness of the collaborative argumentation CALL environment, which may foster a setting where university students can generate reasoning artifacts of higher quality. While the current study has generated outcomes that bear significance for both theory and practice, it is essential to acknowledge the limitations inherent in its design. Primarily, the study's sample size of 55 participants restricts the extent of its contribution to the existing research literature. Consequently, the study's results and implications should be viewed as preliminary, tentative, and exploratory, rather than yielding a definitive conclusion. In order to validate the effectiveness of the collaborative argumentation environment, more samples may be needed to provide a deeper analysis. Additionally, the analysis did not incorporate an examination of the patterns of informal reasoning as the past studies did (Sadler & Zeidler, 2005). Consequently, future research endeavors should incorporate classroom observation as part of the data collection process, as previous studies have indicated that such an approach can provide more extensive and meaningful data (Dawson & Carson, 2020; Dawson & Venville, 2010).

5. Conclusions

In conclusion, this study represented preliminary research aimed at enhancing the informal reasoning abilities of EFL university students. This objective was achieved by implementing a collaborative argumentation CALL environment. The results of this study demonstrated that our university students, who utilized the collaborative argumentation CALL environment, exhibited superior quality of informal reasoning skills compared to those in a traditional argumentation environment. It can be inferred that this CALL environment, Knowledge Forum, facilitated the generation of a greater number of arguments, including counterarguments and rebuttals, among students. These findings hold potential for the application of collaborative argumentation CALL environments in the field. Further research is warranted to investigate the efficacy of different teaching contexts or methods.

Acknowledgements

Funding of this research work was supported by the National Science and Technology Council, Taiwan, under grant number NSTC 111-2410-H-A49-024.

References

Azimah A. A. & Marlizayati, J. (2023). The effect of argumentation about Socio-Scientific Issues on secondary students' reasoning pattern and quality. Research in Science Education, 53, 771-789 (2023). https://doi.org/10.1007/s11165-023-10099-5

Kathpalia, S. S., & See, E. K. (2016). Improving argumentation through student blogs. *System*, *58*, 25–36. https://doi.org/10.1016/j.system.2016.03.002

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536. https://doi.org/10.1002/tea.20009

Scardamalia, M. (2004). CSILE/Knowledge Forum®. *In Education and technology: An encyclopedia*, 183-192. Santa Barbara: ABC-CLIO.

Toulmin, S. (1958). The uses of argument. *Cambridge University Press*.

Wu, Y. T. (2013). University students' knowledge structures and informal reasoning on the use of genetically modifiedfoods: Multidimensional analyses. Research in Science Education, 43, 1873–1890.

Wu, Y. T., & Tsai, C. C. (2007). High school students' informal reasoning on a socio-scientifc issue: Qualitative and quantitative analyses. *International Journal of Science Education*, *29*(9), 1163–1187. https://doi.org/10.1080/09500690601083375.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16926

Collegial evaluation of online English for Specific Purposes (ESP) courses

Christopher Allen[a] and Maria del Carmen Boloña[b]

[a]Department of Language, Linnaeus University, christopher.allen@lnu.se and [b]Universidad Casa Grande, mbolona@casagrande.edu.ec

How to cite: Allen, Ch.; del Carmen Boloña, M. (2023). Collegial evaluation of online ESP courses. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16926

Abstract

This paper reports on a joint pilot study project between two universities in Ecuador and Sweden to develop a practical working framework for the evaluation of each respective institution's online/blended courses in English for Specific Purposes (ESP). The basis for the evaluation is the Conversational Framework (Laurilland, 2012), later developed in the form of a MOOC course in online and blended learning. This course is offered by the Future Learn social learning platform, which offers a large variety of online courses from a consortium of universities worldwide. The teaching and learning of ESP is characterised as the development of learner concepts and practice through interaction between the instructor and learner peers through collaboration and interaction. The learning process in ESP is envisaged using the Conversational Framework in terms of six basic learning types: acquisition, collaboration, discussion, inquiry/investigation, practice, and production. Our work reports on the process of assessing each other's online courses in terms of the extent to which opportunities are provided for students to engage in these learning types. Results from this pilot study suggest that the Conversational Framework can provide a simple, robust, and transparent basis for the initial evaluation of online courses.

Keywords: *ESP, teacher assessment, collaborative work, Conversational Framework, online and hybrid learning environment.*

1. Introduction

The aim of this paper is to examine the efficacy of using a well-established course design framework as the basis for institutional colleagues to compare and benchmark online courses in English. The purpose is to work towards improvements in quality assurance and harmonization of design standards. Our work concerns a joint project between Linnaeus University in Sweden and Casa Grande University in Ecuador during Spring 2023. We follow Lindqvist (2019, p. 505) in referring to collegial learning as, "*learning by means of teachers teaching other teachers within work groups, programs, institutions, and between universities*." The project makes extensive use of Laurilland's (2012) *Conversational Framework.* In our case study we use this Conversational Framework to review each other's online/blended *Moodle* courses. From the Swedish side, the course reviewed was the blended *English for Administrators* course. The Ecuadorean course was called *Contemporary Society: Conflicts and Consensus,* designed as a B1-B2 proficiency level content course in English. In this course, students were presented with course content in the form of contemporary issues in the realms of international politics and global themes.

Within both Linnaeus and Casa Grande universitities respectively, faculty deans have pointed towards the need to harmonise online content around a core of learning activities and structural/navigational commonalities. Drawing

on recent research and initiatives in digital literacy (e.g. Pegrum *et al.*, 2022), there is an important faculty agenda for teachers to develop and update their learning design skills.

The Conversational Framework (Laurilland, 2012) offers an accessible outline for the practical design of online and blended courses within the contexts of formal higher education, using a VLE such as *Moodle*, *Canvas*, or *It's Learning*, etc depending on the institution. The six types of learning identified in this Conversational Framework are included below. Each learning type, with examples using conventional and digital technology, are also presented here in condensed form. The table has been adapted from Laurilland (2012, p. 96):

Table 1. Laurilland's (2012) Conversational Framework learning types with conventional and digital examples.

Learning through	Example: Conventional technology	Example: Digital technology
Acquisition	Reading books, listening to teacher presentations	Reading multimedia, website information
Inquiry	Using text-based study guides	Using online advice and guidance
Practice	Practising exercises and projects, field trips	Using models, simulations, virtual labs
Production	Producing statements, essays, reports	Producing and storing digital documents, e-portfolios
Discussion	Tutorials, seminars, classroom discussions	Online tutorials, seminars, forum discussion threads
Collaboration	Small group projects, discussing others' outputs	Wiki collaborative writing projects, chat

It is suggested that a successful online or blended course should offer participating students as many of these six activities: *acquisition, inquiry, practice, collaboration, and production* (Table 1) as possible, although in practice it might be somewhat difficult to encapsulate all activities within a single course. Conventional activities might be thought of as more traditional oral discussion or 'pen and paper'/print activities which do not require any form of digital tool, resource, or environment. Digital activities in contrast make use of computer hard/software and digital tools and resources such as wikis, online forums, and chat rooms, etc.

When learning through the first of the six activities, *acquisition*, the framework envisages learners passively listening to an online lecture, reading a book, etc. *Inquiry* activities involve learners actively formulating a question which they seek answers to from a library or the internet among course participants. Alternatively, learners can develop their understanding of concepts through the learning type *discussion*, involving the exchange of information and ideas in discussions amongst each other. In learning through *practice*, learners generate an action in response to course goals which are then refined through practice. *Collaborative* activities involve the learners coming to an agreement in producing a shared output. Finally, in learning through *production*, the shared/agreed output from the learners is evaluated by the teacher. The claim is that the six learning types offer a practical, concrete, tangible taxonomy of learning activities to permit colleagues to quickly and easily offer feedback on each other's courses prior to their delivery or following a revision of the course on the basis of course evaluations from previous student groups.

2. Method

2.1 General comments on methodology

The project can best be seen as having closest affinities with a comparative case study given the limited focus on two *Moodle* courses designed and taught by the Swedish and Ecuadorean instructors respectively. The use of the term '*case study'* might be defined as the study of the *"particularity and complexity of a single case"* (Stake 1995, p, xi). In this project, however, the study is effectively a comparison of *two* cases using the Laurilland's (2012) Conversational Framework as a comparative basis. Case studies are seen in this sense as a means of comparing two complex institutional virtual learning environment contexts with "clearly defined boundaries" in Dörnyei's (2007, p. 151) terms. Following Stake (1995, 2005), such a study constitutes an intrinsic case study of the inherent qualities and attributes of the phenomena under consideration.

As with all comparative case studies, one of the main considerations is the generalization issue - the extent to which idiosyncratic features of the courses concerned can be generalized for the purposes of incorporation into a faculty checklist of learning activities (Dörnyei 2007, p152). In this sense both courses being compared here in their Swedish and Ecuadorean contexts are very probably unique. The use of the Conversational Framework with its more general learning type categories can serve as the basis for making generalizations useful to all instructors and course designers of online and blended courses. Table 2 below sets out a brief overview of the two courses.

Table 2. Comparative overview of courses.

Course	English for Administrators	Contemporary Society
Instructor	British native speaker, 35 years' teaching experience PhD applied linguistics	Spanish native speaker, c. 40 years' teaching experience MA degree in TEFL
Course type	ESP (standalone, in-service)	ESP (part of BA undergraduate degree in English)
Delivery mode	Blended/hybrid	Online
Course length	3 months	2 months
CEFR[1] level of participants	C1	B1-B2
Content	English for the higher education sector	Contemporary political and global issues discussed and reported on in English
Number of participants	20	30
Number of classroom sessions	6 (blended/hybrid)	-
Number of online sessions	2	9

[1] Common European Framework of Reference levels, A1(beginner) to C2 (advanced)

2.1. Context and participants

The Swedish course *English for Administrators* was a hybrid/blended/online course with 12 hours of face-to-face plus two Zoom meetings (4 hours in total), combining classroom tasks and online activities with submissions of writing tasks and oral video presentations on site. Both the Swedish and Ecuadorean courses were selected on the basis of their specialized content. Course participants were civil servant employees who worked as administrators at Linnaeus University and generally had a high level of proficiency in English (at least C1in the CEFR framework). This course focused on the terminology of higher education in English (in broad terms, including the vocabulary relating to academic titles and degrees, admissions, credit transfer, study financing, and research, etc) and revision of practical grammar and usage applied to formal report writing for the public administrative sector, with a focus on higher education. The course also focused on a revision of contrastive practical grammar to improve the participants' ability to write formal, standard British or American English depending on their preference in a variety of job-related contexts. The *Contemporary Society* course with 30 students at B1-B2 level in English was worth three ECTS equivalent credits with a total of nine timetabled sessions, and further hours of synchronous and asynchronous interaction online involving specialist terminology, discussion of six main topics dealing with current global and societal issues and research projects as a platform to master content, to become aware of local/global problems and propose solutions. Synchronous content involved instructor-led discussions and presentation of language proficiency issues arising from the discussions while the asynchronous course activity involved reading of texts, watching external video content related to the themes, and forum postings.

2.2. Collegial reviewing

The instructors met online using the video conferencing tool *Zoom* on one occasion in June 2023 for approximately one hour (Figure 1).

Figure 1. Screenshot of the Zoom session Part 1 (of 4 sessions) with instructors comparing the courses.

The tool allowed the sharing of screens, enabling each author to describe in detail the course outline as it is presented on the *Moodle* site. This session was recorded in four parts to provide a permanent record of the discussion and comparison; approximately 30 minutes each of presentation and then discussion was devoted to

each respective course site, totalling one hour. During the course of the presentation, each author went through a protocol based closely on the Laurilland (2012) learning types as defined in the Conversational Framework.

3. Results and discussion

The results of the review are included in Table 3 below. For the *English for Administrators* course, the examples are taken at random from one particular topic on practical grammar and writing, focusing on present simple vs. present continuous tense confusions which can be significant for some Swedish writers of English. The examples from the *Contemporary Society* course are taken once again randomly from the separate topic areas of *Language and Us, Wealth, Gender* and *Environment* issues used as focal points for fostering oral and written forum discussion in English.

Table 3. A comparison of the two courses using Laurilland's (2012) Conversational Framework of six learning type activities.

Activity	English for Administrators	Contemporary Society: Conflicts and Consensus
Acquisition	accessing information about uses and forms of present simple and present continuous in a PPT.	accessing information in the Unit I *Moodle* book, introductory material to activate existing knowledge of topic
Inquiry	identifying simple present and present continuous examples in a PPT taken from official university documents	watching Mark Pagel lecture (*TED Talk*) and main notes to identify salient points in the talk
Practice	analysis of form and function for simple and continuous aspect in example sentences taken from the Swedish higher education sector and practising both tenses in oral interaction	creating a glossary of key terms; identifying facts from opinions in Mark Pagel's lecture
Discussion	in talking about their workplaces, course participants used tense forms to talk about the advantages and disadvantages of working from home/on campus	course participants discussed Mark Page´s proposal, *One World*. One language in breakout rooms and in an online class forum
Collaboration	based on PPT presentations in *Moodle*, course participants collaborated in *Moodle* wikis and in video recordings to describe future challenges facing the university	Sharing opinions about the question, "Are language and culture separable in an opinion blog?"
Production	course participants produced example sentences contrasting simple and continous tenses to describe their job as university administrators. Task of job description submitted for evaluation by the teacher	writing vocabulary definitions, publishing research project findings in a collaborative wiki, including idenfification of a problem, a demographic group affected by this problem, the causes of the problem, one effective way of solving the problem, an approximate overall cost of the solution and how the solution will be implemented.

Table 3 above gives examples of digital learning tasks from both courses. Both courses have a focus on the learning category of *production* involving teacher evaluation of shared output. For the *Contemporary Society* course, this is reflected in the use of the *glossary Moodle* tool where students are tasked with writing definitions of vocabulary items in the reading texts. The defining of vocabulary items using simpler English is similar to a dictionary's use of a defining vocabulary and may be considered as a good example of an activity drawing from CLT as opposed to translation. The module makes good use of writing summaries from online sources such as *Ted Talks*, which is an important academic skill. The module also involves the writing of blogs and collaborative writing activities using the *Moodle* wiki.

4. Conclusions

Our initial reflections point towards the efficacy of using Laurilland's (2012) Conversational Framework as a 'surface' checklist of learning activities which teachers can use for collegial reviewing of online and blended courses. Although broadly coming under the heading of ESP courses, the courses reviewed differed in terms of educational context, content, level and background of students, and course aims. Nevertheless the activities for both courses could be accommodated within the framework. It is obviously difficult to define what is meant by a 'good' course purely in terms of the Conversational Framework's six categories. A more detailed answer to this question is beyond the scope of this paper and would presumably involve the question of whether there is a convergence between languages skills linguistic and specialist content criteria and the course syllabi respectively. Furthermore, the framework says nothing about other important aspects of successful ESP Moodle course design, such as site navigation and logical structuring of course elements either on the basis of thematic or chronological ordering. Nevertheless, we feel encouraged from the initial results of this pilot study to recommend the framework as a basis for collegially developing online courses in our respective institutions.

Acknowledgements

We would like to thank the Department of Languages, Linnaeus University for support in attending the EUROCALL 2023 conference.

References

Buckminster Fuller, R. (1957). *A Comprehensive Anticipatory Design Science*. *Royal Architectural Institute of Canada.* 34. *Retrieved 2023-11-09* – via Google Books.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. UK: Oxford University Press.

Lindqvist, M.H. (2019). The Uptake and Use of Digital Technologies and Professional Development:Exploring the University Teacheer Perspective. In A. Elçi, L. Beith, & A. Elçi (Eds.), *Handbook of Research on Faculty Development for Digital Teaching and Learning* (pp. 505-525). IGI Global.

Laurilland, D. (2012). *Teaching as a design science: Buidling pedagogical patterns for learning and technology*. RoutledgeePegrum, M., Hockly, N. and Dudeney, G. (2022) *Digital Literacies* 2nd Edition. Routledge.

Stake, R.(1995). *The Art of the Case Study Research.* Thousand Oaks, US, California: Sage.

Stake, R. (2005). Qualitative case studies. In N. K. Denzin & Y.S.Lincoln (Eds). *The Sage Handbook of Qualitative Research*, 3rd Edition. Thousand Oaks, California: Sage.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16968

How do virtual reality lessons affect EFL learners' perceptions of speaking English?

Yoshiho Satake[a], Shinji Yamamoto[b] and Hiroyuki Obari[c]

[a]College of Economics, Aoyama Gakuin University, t31330@aoyamagakuin.jp; [b]College of Economics, Aoyama Gakuin University, sya@aoyamagakuin.jp and [c]College of Economics, Aoyama Gakuin University, obari119@gmail.com

How to cite: Satake, Y.; Yamamoto, S.; Obari, H. (2023). How do virtual reality lessons affect EFL learners' perceptions of speaking English? In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16968

Abstract

This study investigates the effects of fully immersive Virtual Reality (VR) lessons on Japanese English learners' perceptions of speaking English. Seven undergraduate students volunteered for one-hour interviews after participating in VR English conversation lessons over seven months using Oculus Quest 2 VR headsets and the Immerse platform. Participants reported a favorable acceptance of VR for English-language learning because of its fun and comfortable nature compared to face-to-face or Zoom lessons. They felt more comfortable and less anxious when communicating in English through avatars and appreciated the gaming aspects of the VR experience. A positive shift was found in the attitudes of those who had not previously experienced English conversation lessons. However, participants highlighted the need for technical improvements in VR technology. These findings suggest VR's potential as an effective language-learning tool.

Keywords: *Virtual Reality (VR), English language learning, foreign language anxiety, avatar.*

1. Introduction

Since the onset of the COVID-19 pandemic, Virtual Reality (VR) tools have emerged as a viable option for remote learning. When applied to language education, these tools are claimed to facilitate learning through communication in virtual contexts (e.g. Chen et al., 2022; Zheng et al., 2022). VR-based communication with avatars can alleviate learners' psychological burden when speaking English and mitigate foreign-language anxiety (Melchor-Couto, 2017). In their research, the authors investigated the effects of fully immersive VR on Japanese English learners' conversational skills, anxiety and confidence levels during English conversations, and intercultural sensitivity. The results suggest that VR reduced anxiety toward English conversation and improved participants' confidence (Satake et al., forthcoming). Therefore, this study explores the reasons behind these observed effects.

1.1 Literature review

VR is increasingly recognized as an innovative tool for pedagogical purposes, particularly in language education. Ma and Zheng (2011) identified three distinct levels of immersion within VR environments: non-, semi-, and fully immersive. Non-immersive environments are typically computer-based, using conventional input methods

such as a keyboard and mouse and limited display capabilities. Semi-immersive VR uses larger screens and may incorporate gesture recognition for more organic interaction, whereas fully immersive systems employ head-mounted displays to create a compelling sense of total immersion.

In an analysis of 26 studies published from 2015 to 2018, Parmaxi (2020) found that primarily non- and semi-immersive VR applications enhanced interaction authenticity and cultural learning and reduced anxiety related to foreign-language learning. However, Parmaxi emphasized the need for educators and experts to thoroughly comprehend VR's potential benefits and applications within the language-education context before widespread deployment.

Dhimolea et al. (2022) examined 32 scholarly articles published between 2015 and 2020, primarily focusing on non- and semi-immersive VR in language education. They concluded that repeated exposure to VR was crucial for successful learning, with the technology proving particularly beneficial for vocabulary acquisition, and that learners' attitudes toward language learning within VR environments were generally positive.

In our previous study (Satake et al., forthcoming), we investigated the effects of fully immersive VR on English conversational abilities, emotional responses, anxiety levels, self-confidence during English conversation, and cultural sensitivity of 102 intermediate-level Japanese English learners (59 experimental and 43 control). Our findings did not reveal statistically significant differences in spoken-language test scores between groups or between the pre- and post-test results. However, we identified a notable enhancement with a substantial effect size in the questionnaire responses. There was no significant change in cross-cultural sensitivity. Students reported that interacting in English with VR avatars was less stressful and more comfortable than in-person English interactions. The research thus supports that the instructional application of VR in English-speaking classes can potentially diminish students' anxiety and boost their confidence in English-speaking abilities.

Although we previously confirmed reduced foreign-language anxiety and increased confidence among Japanese learners in VR English conversation lessons, the reasons remain unclear. Accordingly, this study interviewed participants in the earlier study regarding these effects.

2. Method

2.1. Participants

Figure 1. Screenshot of the Immerse platform.

Seven individuals were selected from the original pool of 59 participants who engaged in VR English conversation lessons as part of our previous study. They received a comprehensive explanation of this study and agreed to participate. The authors' previous research involved participants using Oculus Quest 2 VR headsets and the Immerse platform for English lessons (see Figure 1) several times a month over seven months, featuring avatar-based interactions (Satake et al., Forthcoming). Participants are identified using alphabetical designations (e.g. Student A, B) to ensure anonymity.

2.2. Interview

Two of the authors conducted one-hour Zoom interviews in Japanese with the seven participants, organizing them in pairs or trios. We asked all the participants the same questions based on a pre-established questionnaire. When the participants did not fully answer a question, we moved to the next question without pressing them further. The questionnaire comprised the following items:

1. Have you ever attended non-VR English conversation lessons?

① If yes, please describe the lessons (including your age, frequency, duration, and whether they were conducted face-to-face). Could you share what you liked and disliked about the lessons?

② If no, could you explain why you have not considered attending such lessons until now?

2. What aspects did you like and dislike about the VR English conversation lessons?

3. Has your perception of English conversation changed since your experience with VR lessons?

4. Are there any areas in VR English conversation lessons that need improvement?

The authors transcribed the recorded interviews and created a corpus of 1787 words from 109 responses. This corpus underwent morphological analysis using KHCoder, a quantitative text-analysis software, to separate the words for subsequent cluster analysis. In Japanese, words are written without spaces, necessitating this approach. Cluster analysis, a statistical technique that groups similar items into clusters, provided an overview of the participants' statements. Finally, each response was assigned the number of the respective question, and qualitative analysis was conducted.

3. Results and discussion

3.1. Overview

Figure 2 illustrates the results of the cluster analysis. From left, the first cluster comprises words such as 印象 'impression', 見る 'see', 実際 'actual', 海外' overseas', 面白い 'interesting', アバター 'avatar', 対面 'face-to-face', and 良い 'good'; the second, 自分 'I, me', 空間 'space', 話せる 'able to speak', 多い 'often', 英語 'English', and 伝わる 'understood'; the third, *VR*, 意味 'meaning', *Zoom*, 授業 'class', 英会話 'English conversation', 文法 'grammar', 行く 'go', and 慣れる 'get used to'; the fourth, 先生 'teacher', 楽しい 'fun', 受ける 'take', 意識' awareness', and 言う 'say'; the fifth, 難しい 'difficult', 家 'home', 本当に 'really', ゲーム 'game', 感覚 'feeling', 変わる 'change', and 声 'voice'; the sixth, メリット 'advantage', 内容 'content', 知る 'know', 相手' other people', and 顔 'face'; and the seventh, 向こう 'companion', 話す'speak', ある程度 'to some extent', 会話 'conversation', 楽 'comfortable', 表情 'expression', and 見える 'see'. Referring to the interview corpus and the words' context, the results of the cluster analysis can be summarized as follows:

While face-to-face conversation is good, seeing overseas as an avatar in a virtual space was interesting. Their English was often well understood when they spoke in the virtual space. Being accustomed to VR, they could speak in English conversation classes despite not being good at grammar. VR lessons were better than Zoom lessons, with fun classes. Though operating VR equipment at home was difficult, it felt like a game, and they did not feel the need to change their voice. Not showing their face to others was advantageous, and conversations were comfortable when companions did not see their facial expressions. The participants' assessments of the VR English conversation lessons were favorable.

While Chen and Kent (2020) suggested that learners could speak more freely and comfortably when anonymized by avatars, this study indicates that even when learners were aware of each other's identities, the inability to see facial expressions put them more at ease during conversations. This may reflect a characteristic caution among Japanese learners in communication settings.

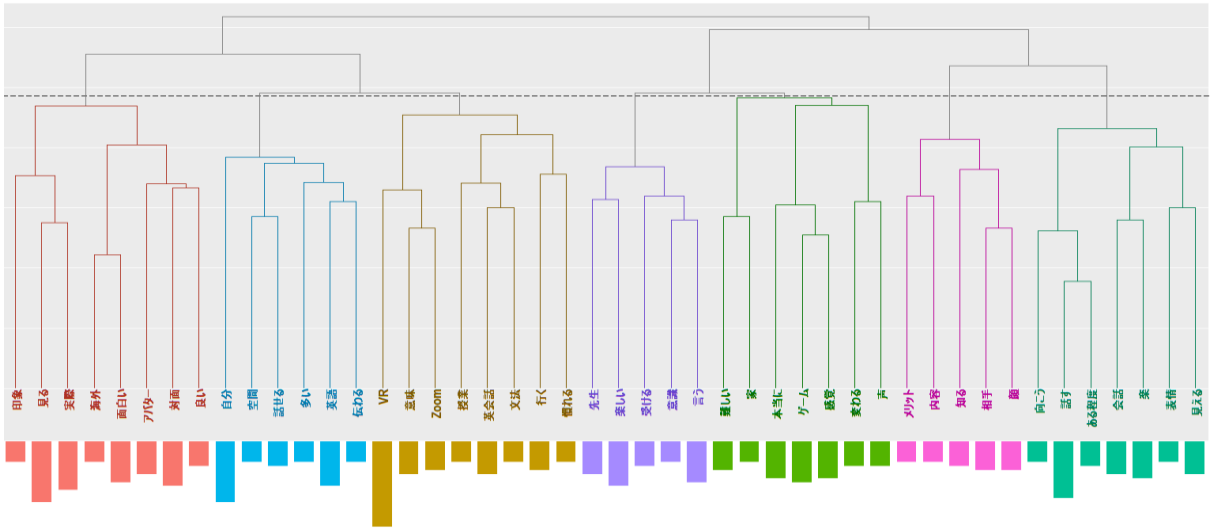


Figure 2. Cluster analysis of the interviews.

3.2. Experience in English conversation lessons

Three of the seven participants (Students A, C, and D) had previous experience with English conversation lessons. Students C (who had weekly lessons from age five to elementary school) and D (who had lessons during kindergarten) had face-to-face English conversation lessons in their childhood, while Student A had online English conversation lessons four times a week after entering university. Student C spoke about enjoyable lessons, such as playing games, and appreciated seeing people's facial expressions in face-to-face lessons. Those with experience in English conversation lessons generally had less foreign-language anxiety and therefore positive feelings about it. The other four participants (Students B, E, F, and G) had no prior experience with English conversation lessons. Student E believed that English should be spoken with grammatical perfection, a sentiment that reflected some foreign-language anxiety. Generally, when this study commenced, those with previous experience held positive feelings about English conversation, while those without experience had somewhat negative feelings.

3.3. Assessment of the VR English conversation lessons

Responses regarding the participants' assessment of the VR lessons comprised 1607 words, accounting for 89.2% of the responses and making it the most talked-about subject, about which participants were most interested in expressing their opinions. Because nouns indicated the topics and the participants' assessment was reflected in adjectives and adjectival verbs, we focused on the high-frequency nouns, adjectives, and adjectival verbs that the participants used. Table 1 shows the nouns, adjectives, and adjectival verbs that appeared more than five times in the assessment section. The nouns imply that the participants were interested in discussing topics related to VR English conversation, their conversation partner and/or teacher, face-to-face and Zoom lessons, feeling like playing a game, and using avatars. The high-frequency adjectives and adjectival verbs suggest that the participants positively assessed the VR English lessons. Referring to the context before and after the high-frequency words, we can summarize the participants' assessment: They enjoyed visiting various places abroad in the VR space, having fun playing sports, and engaging in game-like activities with the teacher and other participants. Avatars hid their faces from their conversation partners, allowing them to converse in English

more comfortably. Of the English conversation methods, face-to-face was considered the best, followed by VR. They would prefer VR over online Zoom lessons if the cost were the same.

Table 1. High-frequency nouns, adjectives, and adjectival verbs in the assessment section.

Noun	Frequency	Adjective/Adjectival verbs	Frequency
VR	21	楽しい' fun'	11
(英)会話 '(English) conversation'	16	面白い 'interesting'	10
相手・向こう' other people/companion'	12	楽だ 'comfortable'	9
英語 'English'	11	良い 'good'	6
対面 'face-to-face'	11		
ゲーム 'game'	10		
先生 'teacher'	8		
アバター 'avatar'	7		
Zoom	7		

Among the high-frequency words, 'game' and 'avatar' distinguished VR English conversation lessons from other English conversation lessons. Participants pointed out that the game-like aspect made it easy to continue conversations without overthinking about what to say, a common occurrence in face-to-face or Zoom online lessons. Regarding the use of avatars, participants found it comfortable not to have their facial expressions revealed, to read their partner's facial expressions, or to physically prepare for the lesson.

One participant appreciated using an avatar because it allowed them to avoid feeling awkward when experiencing anxiety while speaking English. Some participants mentioned that not needing to interpret the teacher's facial expressions made them feel more comfortable while speaking in English because they did not have to consider the teacher's feelings. The female participants especially highlighted the absence of a need to physically prepare. Nevertheless, there was a contradiction in that face-to-face lessons, in which expressions are visible, were still considered the best mode of lessons, and the interviews did not yield any convincing reasons for this preference.

3.4. Changes in impressions of speaking English

No significant change was found in the attitude toward speaking English among the participants with prior experience with English conversation lessons. However, students without such experience reported a positive shift in their perception. They found VR-based English conversation lessons fun, enabling them to speak even without perfect grammar. Moreover, their fear of speaking in English diminished after these VR lessons. Briefly, positive feelings about speaking English did not change among those who had experienced English conversation lessons, but they changed from negative to positive among those without such experience.

3.5. What needs to be improved in VR English conversation lessons?

The participants' feedback indicates several areas in VR technology that require improvement, including audio issues, such as: difficulties with hearing; space-related challenges when using VR at home and during

movement; poor responsiveness of movement controls to match users' desired actions; the complicated setup process; and discomfort, such as headaches from the weight of the VR goggles. Furthermore, the participants suggested the option to choose between avatars and real faces, depending on their preference to read facial expressions. Additionally, they recommended personalizing the avatars to better reflect individual characteristics. In short, the participants' feedback suggests a need for various improvements in VR technology.

4. Conclusions

Our study suggests that VR English conversation lessons can positively impact Japanese English learners by creating a comfortable learning environment. The game-like nature of VR and avatars eases the pressure associated with face-to-face interactions, thereby reducing foreign-language anxiety. While participants with no prior English conversation experience reported a positive attitude shift toward speaking English, several technological improvements were suggested to enhance the VR learning experience. Future research should aim to optimize the use of VR in language learning and address its limitations.

Acknowledgments

This study was financially supported by the College of Economics, Aoyama Gakuin University.

References

Chen, B., Wu, Y., & Wu, L. (2022). The effects of virtual reality-assisted language learning: A meta-analysis. *Sustainability*, *14*(6), 31-47. https://doi.org/10.3390/su14063147

Chen, J. C., & Kent, S. (2020). Task engagement, learner motivation, and avatar identities of struggling English language learners in the 3D virtual world. *System*, *88*, 102168. https://doi.org/10.1016/j.system.2019.102168

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, *66*, 810-824.

Ma, M., & Zheng, H. (2011). Virtual reality and serious games in healthcare. In S. Brahnam & L. C. Jain (Eds.), *Advanced computational intelligence paradigms in healthcare 6: Virtual reality in psychotherapy, rehabilitation, and assessment* (pp. 169-192). Cham: Springer. https://doi.org/10.1007/978-3-642-17824-5_9

Melchor-Couto, S. (2017). Foreign language anxiety levels in Second Life oral interaction. *ReCALL*, *29*(1), 99-119. https://doi.org/10.1017/S0958344016000185

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments*, *31*(3), 172-184. https://doi.org/10.1080/10494820.2020.1765392

Satake, Y., Yamamoto, S., & Obari, H. (Forthcoming). Effects of English-speaking lessons in virtual reality on EFL learners' confidence and anxiety. Abingdon: Routledge.

Zheng, C., Yu, M., Guo, Z., Liu, H., Gao, M., & Chai, C. (2022). Review of the application of virtual reality in language education from 2010 to 2020. *Journal of China Computer-Assisted Language Learning*, *2*(2), 299-335. https://doi.org/10.1515/jccall-2022-0014

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16955

App-assisted language revitalization: Insights from applied cognitive linguistics

Christina Ringel[a] and Theresa Pohle[b]

[a]DoKoLL, TU Dortmund University, christina.ringel@tu-dortmund.de and [b]Freelance software developer, linguist, theresa.freelance@proton.me

How to cite: Ringel, C.; Pohle, T. (2023). Basing app-assisted language revitalization on applied cognitive linguistics. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16955

Abstract

Current theoretical advances in applied linguistics have not yet found wide practical application in the field of language revitalization. In this paper, plans for an open source application for desktop computers and mobile devices for Indigenous language learning settings will be outlined. The app consists of building blocks inspired by cognitive linguistics and task-based language learning. Members of Indigenous language communities can use these to create exercises and assessment modules for their respective languages. In the paper, a mock-up with model exercises will be showcased to illustrate how certain aspects of the afore-mentioned theories can be applied. For example, vocabulary tasks are informed by insights from the analysis of collocations, connotations, frames, metaphors, prototypicality, and semantic relations.

Keywords: *computer/mobile-assisted language learning, task-based language teaching, cognitive linguistics, endangered languages.*

1. Introduction

In many endangered language communities around the world, continuous hard work has been undertaken to document and revitalize their languages and cultures. For instance, traditional documentation and revitalization efforts include the development of writing systems, the production of dictionaries and grammars, publications and media productions, language classes and programs such as Language Nests (Grenoble & Whaley, 2006), and Master-Apprentice Language Learning Programs (MALLP; Hinton et al., 2018).

Considering these programs, it is noticeable that certain findings in the field of applied linguistics have not yet found wide practical application, in particular Cognitive-Linguistic (henceforth CL) approaches as well as task-based language learning. Based on these findings, we have developed a project plan to create an application for Indigenous language learning settings that will put current theory into practice. In this paper the application's fundamental principles to date will be presented and we will delve into the aspects and applications of the aforementioned theories that will form the basis of the app user's learning experience.

2. Project outline

2.1. Theoretical foundation

Since the early 1990s, researchers have conducted research at the cognitive-didactic crossroads (for instance, Dirven, 1989; Taylor, 1993). In the context of Indigenous languages, Ahlers (1999) successfully explored this field to enhance the Hupa community's revitalization efforts. Her understanding is that

> language is seen as natural outcome of humans…, who have exploited their rich cognitive resources–such as the ability to focus attention, to automize, to categorize, to form generalizations…–to develop language, that is, to express meaning. … Instead of an emphasis on rules or abstract principles particular to language, CL focuses on established cognitive and perceptual principles…. Such a view of language entails that morphosyntactic patterns are not meaningless, as they reflect, albeit in rather abstract form, human conceptualizations (Verspoor & Tyler, 2007, p.160).

Much of the pioneer work of early cognitive linguists was, in fact, done with Indigenous languages. As Rice puts it, "some of the core tenets of the field are based on analysis and promotion of actual usage, the ubiquity of metaphor and metonymy in lexicalization and constructionalization, and the primacy of situated and embodied interaction" (Rice, 2019a). While, according to Rice, this is exactly what has

> long provoked a high degree of intellectual excitement and wonderment at the diversity and 'special genius' of individual languages [in researchers], speakers and learners of threatened languages can come to appreciate this same excitement… when helped to approach their languages intuitively and from the perspective of meaning and usage – as advocated by CL (Rice, 2019b, p.93).

Additionally, though in more general terms, Littlemore points out that CL "suggest[s] ways in which the relationships between grammatical expressions and their original lexical meanings can be made apparent in the language classroom to enhance learning and memorization" (Littlemore, 2009, p.3).

Another valuable strand of research that will be incorporated in the development of the app is task-based language teaching. Nunan (2004, p.6) presents this methodology by establishing the following principles:

- needs-based approach to content selection

- emphasis on learning to communicate through interaction in the target language

- introduction of authentic texts into the learning situation

- provision of opportunities for learners to focus not only on language but also on the learning - process itself

- enhancement of the learner's own personal experiences as important contributing elements to classroom learning

- linking of classroom language learning with language use outside the classroom

Like CL, task-based language learning allows for insights into language as well as thought processes. On top of that, it ensures that language learning adequately prepares for successful communication in the target language by focusing on task-based and needs-based grammar and vocabulary. In our view, this must be the ultimate goal of any language revitalization activity.

2.2. Current state of app development

While some technical details of the app are yet to be determined, mock-ups of tasks have been created in order to demonstrate some aspects and applications of cognitive linguistics and task-based language learning that will form the basis of the app user's learning experience. One of these mock-up tasks is shown below in Section 2.3.

For example, vocabulary tasks in our app will aim to present lexical items within their semantic networks by providing information about prototypicality, collocations or frames. Furthermore, insights into the motivated rather than arbitrary nature of lexemes will be given by presenting them in the light of conceptual metaphors, embodiment, onomatopoeia or phonesthemes. Similarly, grammatical tasks that will teach the system and underlying constructions of the respective language will be prepared according to suggestions in the relevant cognitive linguistic literature, where available.

2.3. Exemplary vocabulary task

In the mock-up in Figure 1a and 1b, the learner takes part in a cooking session which introduces him or her to the semantic frame of preparing food. The first screen (Figure 1a) depicts an introductory video which shows two people making a dish. Below the video, transcriptions in the target language are available for the learner to follow along. Optionally, translations in the source language are added via tooltips (see Figure 1a).

On the second screen, the learner is asked to apply the newly gained knowledge. In a first task, the learner has to decide on the correctness of prompts (see Figure 1b). In this task, relevant vocabulary is presented in collocation-enriched and frame-fitting context. The exercises highlight additional collocations (e.g. *to make bread*) and introduce related vocabulary (e.g. *soup*, *flour*). In a second task, he or she is asked to fill out gaps (see Figure 1b). This promotes the repetition of constructions presented in the video.

Figure 1a. Mock-up part I.

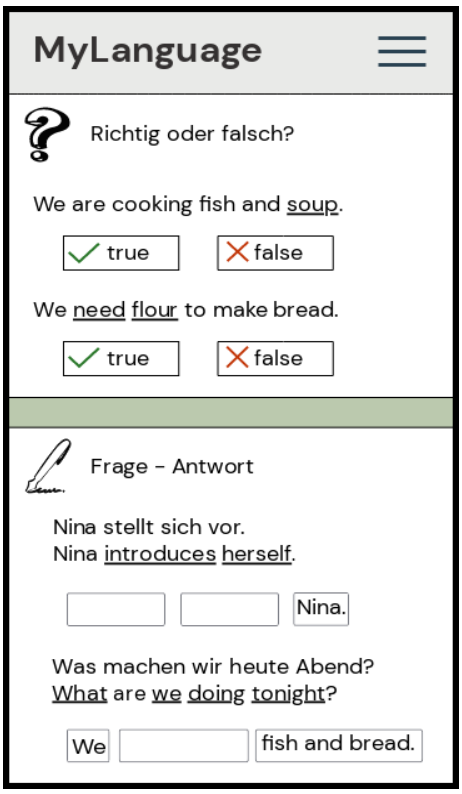


Figure 1b. Mock-up part II

While this mock-up is still at an early stage of development, it serves to showcase a) the future modularity of lesson creation and b) our understanding of task-based learning.

2.4. Ownership

At the core of our idea is the conviction that any tool developed to support language revitalization efforts first and foremost needs to be a gain for the respective language communities. In order to make an impact, tools need to be adaptable to the local context, including the available resources, the goals of the users etc. Further, as Robert Elliott points out, "issues of ownership and control of data and information have historically affected Indigenous and minority communities disproportionately" (2021, p.297).

Thus, in order to guarantee that ownership of language data is exclusive to the community and that community members have maximum control over their language use, we have carefully set parameters for our application which are further detailed in sections 2.5., 2.6. and 2.7.

2.5. Community involvement in design and development

Much like an online platform that allows users to build their own websites by combining building blocks, our app will merely provide a skeleton which local language experts will be able to gradually populate by choosing templates for exercises or assessments and adding content to them. These blocks will be bare and only provide the technical functionality. It will be the community's task to create a curriculum based on their needs and wants, as well as bringing the software to life by adding audio, text, and images.

In order to ensure that the software will be of the greatest use to Indigenous communities, we will be presenting the software prototype to two partner communities during the design and development stages. They will be asked to test and evaluate the software. These meetings with the communities will also involve the first sessions of user training. After the official software release, we will offer software training and technical support for any user to help with potential issues of any kind. We are currently approaching potential partner communities and hoping to

establish ties for collaboration. Furthermore, it is our desire to build a long-term and diverse network of language experts and activists, Indigenous teachers and developers, and Master/PhD students to see our software development continually rooted where it belongs.

2.6. Technological and economic considerations

We are planning to create an application that is both optimized for desktop computers as well as mobile devices. This is to ensure that the majority of users will be able to access it, be it at school, in a language center, on the move, or remotely in the countryside. We also bear in mind that internet access cannot (and should not) always be presumed. The application will therefore be functional both online and offline.

In regards to financing, we are entertaining different options, such as third-party funding, donations, or a very low-level software leasing model. We are also going to take inspiration from the open-source community. The principal aim is to keep the hurdles as low as possible for the communities. Our app will be non-proprietary and ideally free of charge.

2.7. Why yet another app?

The idea to support language revitalization efforts with the help of apps is far from new. Not only the Miriwoong people already have a dictionary and a language learning app at their disposal.[1] What makes our approach special is that (i) it will not be designed for one language in particular, (ii) the infrastructure will be based on theoretical advances in cognitive linguistics and task-based language learning, so that the potential for positive outcomes in language learning is maximized, and (iii) community involvement is central to all stages of the development process.

For example, there will not be a predefined and fixed set of semantic fields to be filled with words from the respective language. Rather, the community can decide which fields are relevant to their culture. Similarly, for language learning exercises, those templates can be selected that best reflect local ways of learning and that can be implemented with the available resources.

3. Goals and Mission

For many Indigenous communities affected by language endangerment, language revitalization is a matter of crucial importance but is also enormously challenging. This process is not merely about teaching vocabulary and grammar. One of its major goals is to promote the community's pride and identity and to strengthen the relationship of the people with country and kinship. Regular language learning technology, as we know it from Western cultures, cannot cater to these manifold and unique aspects.

Our mission is therefore to adopt a new and creative approach which will allow for a learning environment that is, on the one hand, adapted to the human brain, and, on the other hand, rooted in the respective practical context and motivated by local community needs.

Acknowledgements

We would like to thank the Miriwoong community, whose country is located in north-east Western Australia, for inspiring the idea for our app.

[1]An overview of a variety of apps and other digital resources can be found, for example, here: Petersen, Rachael. 2013. iDecolonize: A Review of Indigenous Language-Learning Apps. Retrieved July, 30, 2023, from https://rising.globalvoices.org/blog/2013/06/21/idecolonize-a-review-of-indigenous-language-learning-apps/. The Miriwoong app Learn Miriwoong is described here: Miriwoong Mobile App. Retrieved July, 30, 2023, from http://mirima.org.au/mdwg-work/miriwoong-app/

References

Ahlers, J. (1999). *Proposal for the use of cognitive linguistics in Hupa language revitalization*. [Doctoral dissertation, University of California, Berkley]. Retrieved July, 30, 2023, from https://escholarship.org/uc/item/437086g7

Dirven, R. (1989). Cognitive Linguistics and Pedagogic Grammar. In H. Altmann, H. E. Brekle, H. J. Heringer, C. Rohrer, H. Vater, & O. Werner (Eds), *Reference Grammars and Modern Linguistic Theory* (Linguistische Arbeiten 226) (pp. 56–75). Max Niemeyer. https://doi.org/10.1515/9783111354590.56

Elliott, R. (2021). Technology in Language Revitalization. In J. Olko & J. Sallabank (Eds), *Revitalizing endangered languages: A practical guide* (pp. 297–316). Cambridge University Press. https://doi.org/10.1017/9781108641142.018

Grenoble, L. A., & Whaley, L. J. (2006). *Saving languages: An introduction to language revitalization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615931

Hinton, L., Florey, M., Gessner, S., & Manatowa-Bailey, J. (2018). The Master-Apprentice Language Learning Program. In L. Hinton, L. M. Huss, & G. Roche (Eds), *The Routledge handbook of language revitalization* (pp. 127–136). Routledge. https://doi.org/10.4324/9781315561271-17

Littlemore, J. (2009). *Applying cognitive linguistics to second language learning and teaching*. Palgrave Macmillan. https://doi.org/10.1057/9780230245259

Mirima Council Aboriginal Corporation. (2023, July 31). *Mirima Dawang Woorlab-gerring | Mirima Place for Talking.* http://mirima.org.au

Nunan, D. (2004). *Task-Based Language Teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667336

Petersen, R. (2013, July 31). *iDecolonize. A Review of Indigenous Language-Learning Apps.* https://rising.globalvoices.org/blog/2013/06/21/idecolonize-a-review-of-indigenous-language-learning-apps/

Rice, S. (2019a). Cognitive linguistics and the study of Indigenous languages (Abstract). Retrieved July, 30, 2023 from https://iclc2019.site/wp-content/uploads/abstracts/plenary/rice.pdf

Rice, S. (2019b, August 6-11). Cognitive linguistics and the study of Indigenous languages [Conference presentation]. ICLC-15: Cross-Linguistic Perspectives, Nishinomiya, Japan. https://www.researchgate.net/publication/335109318_Cognitive_Linguistics_and_Indigenous_Language

Ringel, C. Forthcoming. *Rangga ngenandayin, lingbe berranben-nging-ngerri: Possession in Miriwoong, a non-Pama-Nyungan language of north-west Australia*. [Doctoral dissertation, University of Cologne].

Taylor, J. R. (1993). Some pedagogical implications of cognitive linguistics. In R. A. Geiger & B. Rudzka-Ostyn (Eds), *Conceptualizations and Mental Processing in Language* (pp. 201–223). Mouton de Gruyter. https://doi.org/10.1515/9783110857108.201

Verspoor, M., & Tyler, A. (2007). Cognitive linguistics and second language learning. In W. C. Ritchie & T. K. Bhatia (Eds), *The new handbook of second language acquisition* (pp. 160–177). Emerald.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16950

Harnessing the power of images in CALL: AI image generation for context specific visual aids in less commonly taught languages

Liang Xu[a], Elaine Uí Dhonnchadha[b] and Monica Ward[c]
[a]Dublin City University, Dublin, liang.xu6@mail.dcu.ie; [b]Trinity College Dublin, Dublin, uidhonne@tcd.ie; and [c]Dublin City University, Dublin, monica.ward@dcu.ie

How to cite: Xu, L.; Uí Dhonnchadha, E.; Ward, M. (2023). Harnessing the power of images in CALL: AI image generation for context-specific visual aids in less commonly taught languages. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16950

Abstract

This paper explores the application of AI image generation in Computer-Assisted Language Learning (CALL) for Less Commonly Taught Languages (LCTLs). It delves into the potential of text to image generation models in creating context specific visual aids to enhance comprehension and engagement among learners. The integration of AI generated images in a language learning game, Cipher, is discussed, showcasing the benefits and challenges encountered. Learner feedback indicates positive inclinations towards the AI generated images, but also highlights the need for meticulous selection to address biases and stereotypes. Overall, this approach shows promise in creating culturally relevant CALL resources and improving language learning experiences for learners of LCTLs.

Keywords: *CALL, less commonly taught languages, AI image generation, educational games.*

1. Introduction

Images are powerful. With creative ways of harnessing this power, images can be used in education to enhance student engagement. For instance, Callow (2012) presented a pedagogical strategy that leverages students' inherent affinity for images to promote greater participation and involvement in the classroom setting. Research suggests that pictures are more readily identified and retained in memory than words, and visual stimuli are processed at a substantially faster rate compared to textual information (Schroeder et al., 2011). The integration of images and text in Computer-Assisted Language Learning (CALL) can provide numerous advantages, specifically the use of visual imagery can serve as a powerful tool for stimulating learners' interest and engagement, thereby enhancing the overall efficacy of the instructional process.

In the context of CALL for Less Commonly Taught Languages (LCTLs), the incorporation of graphic design elements can be of great benefit. However, due to the relative scarcity of resources available for LCTL CALL development (Ward, 2015) the inclusion of visual design features may be considered a luxury. For instance, in a typical educational game development scenario, the visual elements usually require graphic designers to create imagery based on specific criteria, such as the game's theme, suitability for a certain group of people (e.g. school children), and maintaining cultural relevance. This process is time-consuming and expensive, and many CALL projects focusing on endangered and indigenous languages do not have access to graphic designers due to project

scale and resource limitations. Therefore, we propose AI image generation as a more efficient and cost effective approach to creating context specific imagery in CALL. Image generation tools like Midjourney are able to generate high-quality context specific images based on descriptive prompts.

2. Background: Text-to-Image Generation AI for game assets development

Generative AI, particularly text to image AI has emerged as a powerful tool in game design and the visual arts, enabling the generation of visual content. Text To Image Generation (TTIG) models, such as DALL-E[1], Midjourney[2], Stable Diffusion[3], and Imagen[4], are self-supervised deep learning models trained on massive datasets. They can generate high quality images based on multi modal commands, including text descriptions and user-provided images. These models enable the manifestation of imaginative concepts, fusion of unrelated objects, and variations of existing images (Ko et al., 2023). The game industry has embraced TTIG models (Vimpari et al., 2023). with the CEO of Unity, a leading game engine company, stating that generative AI will have a transformative impact on gaming (Koetsier, 2023).

Certain models have been trained on open source datasets and made available online for people to access freely (Ko et al., 2023). Moreover, models like Stable Diffusion are open source but require a subscription for access. Vimpari et al. (2023) emphasise the empowering capabilities of TTIG models, enabling users to generate visually appealing content without requiring conventional artistic expertise, which can be financially out of reach for LCTLs with limited resources.

TTIG models find applications across various visual art domains such as graphic design, UI design, webtoons, digital art, and new media art. However, there are limitations - they tend to generate predictable images, lack support for personalisation, and can restrict creativity when solely relying on text prompts (Ko et al., (2023). There are also efficiency considerations and there are concerns about reinforcing stereotypes, biased or harmful content, copyright issues, and artists' authority and compensation (Vimpari, et al., 2023). Despite these limitations and concerns TTIG has firmly arrived in the game industry and visual art field and is expected to impact the creative process from early ideation to the final product's assets (Vimpari, et al., 2023).

3. Method: TTIG in CALL for LCTLs

Images can provide more context for the learning materials delivered through CALL. For example, context assists in reading, and images can be used to enhance context. In our CALL project, Cipher (Ward et al., 2022), a digital game-based language learning application, we employ AI-generated images as visual aids to reinforce reading comprehension and enhance gameplay.

[1] https://openai.com/dall-e-2

[2] https://www.midjourney.com

[3] https://stability.ai/stablediffusion

[4] https://imagen.research.google

Figure 1. Illustrations used in the Cipher game

In Cipher, traditional stories and fairy tales are presented to the learners in pages, with approximately 40 words per page. The images for these pages were generated using an iterative process. Initial prompts for the images were fed into the Midjourney and reviewed. The prompts were then modified to achieve better results. Once an image was deemed suitable, further related images for other parts of the story were generated until all the required images had been generated. Generating images with the same look and feel was a challenge. The illustration (Figure 1 left) is used in the story 'Hansel and Gretel' and was generated with the prompt "a witch with her back turned to the fire in her kitchen, with mysterious, transparent background". The illustration (Figure 1 right) is used in the Irish traditional story "the salmon of knowledge" and was generated with the prompt "mystical boy touching fish roasting on a spit over large fire". The screenshot displayed in Figure 2 depicts the utilisation of an AI-generated image within the game.

Research and evidence have shown that visual and multi-modal texts are highly effective in captivating and engaging young people across various domains, such as gaming, social media, video creation, and digital text production (Schwienhorst, 2002; Callow, 2012). The utilisation of AI image generation can result in significant improvements to CALL materials, thereby saving time and resources while also reducing costs. The integration of this approach in CALL holds great promise as a powerful educational tool for learners, particularly in the case of LCTL CALL.

Translation:

"There was a big fire under the salmon.

A blister appeared on its skin.

Oh, said Fionn, I've burnt it.

He pressed his thumb on the blister to burst it.

The skin was hot.

He burned his thumb.

He put his thumb in his mouth. Finnéigeas …"

Figure 2. AI generated image in the game along with translation

Furthermore, AI image generation offers the capacity to enhance the game's thematic focus of reconnecting with the spirit of the language by incorporating indigenous and culturally specific elements, thereby fostering a deeper connection with the language's cultural roots (Napier and Whiskeyjack, 2021). In the realm of indigenous mythology and folklore AI image generation has the potential to produce captivating and unique images that can effectively bring stories to life, especially where such imagery may not have been previously available. This is particularly significant for mythological narratives within indigenous cultures, as these stories are often confined to a relatively small global audience. As a result, the circulation of folklore is limited, leading to a scarcity of related imagery. Even when some images are available, they may not be suitable for educational purposes, or may have copyright restrictions.

4. Results and discussion

This section reports on feedback from learners who have used the Cipher game with integrated AI-generated images. Furthermore, we provide guidance advice for other CALL researchers considering the incorporation of AI-generated images into their projects.

4.1. Learner feedback

A user experience investigation, based on a prior study (Xu et al., 2022), focussed on the images incorporated in the game. The participants consisted of individuals aged between 8 and 12 years attending an English medium primary school for boys. A majority of these participants (71%) reported engaging in gaming activities on a daily basis. Following their engagement with the game over two months (once every week, 30 minutes each time) the students were requested to complete a survey. The response categories in the survey were structured using a 5-point Likert scale ranging from 1=very negative to 5=very positive. Responses rated 4 or 5 were categorised as a positive assessment in accordance with the criteria established in Xu et al., (2022).

Of the total 165 responses, 51% expressed a positive inclination towards the story images featured in the game. With respect to the question concerning the extent to which the images contributed to the comprehension of the narrative, 27% of the respondents reported perceiving the story images as facilitative in enhancing their understanding. Additionally, 28% of the responses indicated that the images could potentially assist their comprehension. More than half of the respondents, 53%, expressed that they liked playing the game. A substantial majority of 66% of respondents found the game's approach to learning Irish to be more enjoyable than traditional classroom teaching. Approximately 38% of respondents expressed a positive sentiment toward learning Irish after playing the game and a little over one third of respondents (34%) believed they learned something while playing the game. More detailed information about the survey and its findings can be found in the study (Xu et al., 2023).

4.2. Challenges and suggestions for future work

TTIG AI is currently undergoing rapid development. At the time of writing (July 2023), Google Imagen, a text-to-image diffusion model, is in its beta release phase (Rangwala, 2023). New features which involve extending an image beyond its original borders were implemented in DALL·E in August 2022 (OpenAI, 2022) and Midjourney in June 2023 (Beyer, 2023). Some TTIG models (e.g., DALL·E and Midjourney) have only recently been made available to the public and, at the time of writing, there are almost no regulations currently in place for these tools. TTIG models are part of a larger Artificial Intelligence ecosystem that has been released since 2022 including Generative Artificial Intelligence (GenAI) tools like ChatGPT. Governments and international bodies such as the European Parliament are drafting laws to monitor AI tools. These advancements highlight the dynamic capabilities of image generation while also highlighting certain challenges. The CALL community has continually explored the possibilities offered by new technologies in the language learning process. TTIG AI has the potential to become a powerful tool in the preparation of materials for CALL. It is important to investigate what TTIG AI can provide, particularly as the technology improves, while at all times adhering to national and international regulations.

Figure 3. A witch embedded in the front wall (left) and a boy with a strange hand (right)

Throughout the incorporation of AI generated images into the game we faced certain challenges. Some images produced by TTIG models exhibited biases and stereotypes. Therefore, a meticulous selection process is imperative prior to the inclusion of these images in the game. Considerable efforts are required to generate improved images using TTIG models. Achieving specific outcomes often requires numerous attempts involving adjustments to prompts or variations in the generated images. Additionally, it may be necessary to make minor

adjustments using additional editing tools, as some generated images may possess imperfections such as extra fingers on a human subject, see Figure 2 right. Unusual and unwanted 'blends', i.e., the blending of objects from the prompts, were a common occurrence. For example, a prompt like "a witch standing in front of her house" could result in an image of a witch embedded in the front wall of a house, see Figure 2 left. Also, the representation of females in the generated images tended to depict stereotypical extremes, portraying women as either youthful and provocative or ancient and repulsive. Any prompts using the word 'goddess' were not suitable for a children's game. It is also difficult to generate a series of pictures for a story which have a common look and style. Nevertheless, with more careful use of prompts and the avoidance of problematic prompt terms very good results are possible.

5. Conclusions

This paper explored the use of text-to-image generation models to enhance the development of CALL resources for LCTLs. The positive feedback received from users of the CALL application highlights the potential of AI-generated images in engaging learners and improving comprehension. However, it is crucial to address challenges present in some AI-generated images. There is a need to be aware of the ethical issues with TTIG, particularly the issues of bias and copyright. Continued research and development in this area can further advance the effectiveness and inclusivity of CALL resources, which may ultimately enhance language learning experiences for learners. In summary, although there are challenges in using TTIG models for image generation or ideas drafting, we believe that the benefits make it a worthwhile endeavour.

Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No.18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Beyer, E., (2023). Midjourney 5.2: A guide to the latest tools and features. *NFT Now Podcast* https://nftnow.com/guides/midjourney-5-2-a-guide-to-the-latest-tools-and-features/

Callow, J. (2012). The rules of visual engagement: images as tools for learning. *Screen Education*, (65), 72-79.

Ko, H. K., Park, G., Jeon, H., Jo, J., Kim, J., & Seo, J. (2023, March). Large-scale text-to-image generation models for visual artists' creative works. *In Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 919-933). https://doi.org/10.1145/3581641.3584078

Koetsier, J. (2023). Generative AI will completely transform gaming as we know it. *TechFirst with John Koetsier*. https://johnkoetsier.com/generative-ai-gaming/

Napier, K., & Whiskeyjack, L. (2021). wahkotowin: reconnecting to the spirit of nêhiyawêwin (Cree Language). *Engaged Scholar Journal*, 7(1), 1-24. https://doi.org/10.15402/esj.v7i1.69979

OpenAI. (2022). DALL·E: Introducing outpainting. *Open AI*. https://openai.com/blog/dall-e-introducing-outpainting

Rangwala, A., (2023). What is Google Imagen AI? How to use it?. *Open AI Master*. https://openaimaster.com/google-imagen-ai/

Schroeder, S., Richter, T., McElvany, N., Hachfeld, A., Baumert, J., Schnotz, W., Horz, H., & Ullrich, M. (2011). Teachers' beliefs, instructional behaviors, and students' engagement in learning from texts

with instructional pictures. *Learning and Instruction*, 21(3), 403-415. https://doi.org/10.1016/j.learninstruc.2010.06.001

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

Vimpari, V., Kultima, A., Hämäläinen, P., & Guckelsberger, C. (2023). "An adapt-or-die type of situation": Perception, adoption, and use of Text-To-Image-Generation AI by Game Industry professionals. *ACM Hum.-Comput. Interact. 7, CHI PLAY*. https://doi.org/10.1145/3611025

Ward, M. (2015). CALL and less commonly taught languages: challenges and opportunities. *Research-publishing. net*. http://dx.doi.org/10.14705/rpnet.2015.000391

Ward, M., Xu, L., & Dhonnchadha, E. U. (2022). Game based language learning for Irish: noticing errors while playing. Intelligent CALL, granular systems and learner data: *short papers from EUROCALL 2022*, 380. https://doi.org/10.14705/rpnet.2022.61.1488

Ward, M., Uí Dhonnchadha, E., McGarry J., and Xu, L. (2023). Co-creating CALL content - does it work? Goldilocks compromise or Cruella chaos? Short papers from EUROCALL 2023.

Xu, L., Dhonnchadha, E. U., & Ward, M. (2022, June). User experience study of "Cipher: Faoi Gheasa", a digital educational game for language learning and student engagement. *In Proceedings of the 2nd Workshop on Games Systems* (pp. 5-8). https://doi.org/10.1145/3534085.3534339

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16988

The teacher-in-the-loop: collaboration with the teaching community in the development of an iCALL platform for Irish

Madeleine Comtois[a] and Neasa Ní Chiaráin[b]

[a]School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, comtoism@tcd.ie and [b]School of Linguistic, Speech, and Communication Sciences, Trinity College Dublin, neasa.nichiarain@tcd.ie

How to cite: Comtois, M.; Ní Chiaráin, N. (2023). The teacher-in-the-loop: collaboration with the teaching community in the development of an iCALL platform for Irish. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16988

Abstract

This paper explores the teacher functionality within An Scéalaí, an iCALL platform being developed for Irish. Aiming to foster Irish language learning, the design and development of An Scéalaí has from the outset employed a cooperative approach with educators to optimise its responsiveness to user needs. There is a particular emphasis on corrective feedback at this current stage of development as we posit that administering constructive corrective feedback to learners will be the cornerstone of its pedagogical effectiveness and user satisfaction. The efficacy of iCALL platforms in promoting endangered languages, such as Irish, is contingent upon the active involvement of the associated language community. This work presents the findings of a recent questionnaire conducted as part of the ongoing initiative to engage educators and developers in a collective effort towards language revitalisation. We sought to gain an overview of teachers' experiences teaching Irish and their impressions of the An Scéalaí platform, introducing new functionality they would not have previously encountered. The results give us insights into the current situation of technology use in classrooms (post-Covid). It reinforces the importance of having the teacher-in-the-loop at every stage of iCALL development to ensure useful applications are built for the wider teacher community.

Keywords: *iCALL, corrective feedback, teacher-in-the-loop, Irish language.*

1. Introduction

The effectiveness of iCALL platforms, as well as the extent to which they are adopted, in the teaching and learning of endangered languages hinges on how much the respective language community participates in its evolution. From the foundation of the Republic of Ireland, education has been recognised as a critical factor in preserving and maintaining the Irish language. Speech and language technologies, when encapsulated within a pedagogically robust iCALL framework, have the potential to transform Irish language education, as they can facilitate interactive and personalised learning experiences that cater to the individual needs of the learners.

This paper describes *An Scéalaí*[1] (the Storyteller), an iCALL platform under development at Trinity College Dublin for the teaching/learning of Irish (Ní Chiaráin et al., 2022). It is being developed as part of a larger initiative, ABAIR[2], which develops core speech technologies, TTS and ASR, for the Irish language. The *An*

[1] scealai.abair.ie

[2] www.abair.ie

Scéalaí platform is designed to facilitate the teaching of Irish at both secondary and tertiary levels, as well as to support autonomous learners in their language acquisition journey. The focus of this paper is the importance of collaborative engagement with educators in the design and development process of the platform, in particular the delivery of corrective feedback. The 'teacher-in-the-loop' concept is often used to define this human activity involved in the training and decision-making process of AI and digital systems. This participatory approach is critical to ensure that the platform is practical, effective, and responsive to the needs of end users. This paper presents insights gathered from a recent questionnaire designed to ascertain developmental priorities, contributing to an evolving understanding of the specific needs and preferences of our users.

For endangered languages, speech technology has an especially important role to play as it enables us to place the spoken language as the focal point of all language-learning activities. Most learners have very limited/no exposure to native speakers, and the ability to access native-speaker models of the language is a substantial advantage.

2. Background

2.1. The Irish language

The Irish language, classified as "definitely endangered" by UNESCO (Moseley, 2010), is a Q-Celtic language closely linked to Scottish Gaelic and the extinct Manx. It has distant ties to P-Celtic languages like Welsh, Breton, and Cornish (also extinct). The Celtic languages face significant challenges related to preservation and revitalisation. Irish, however, is Ireland's first national language and an official EU language since 2007. The 2022 census data indicates nearly 1.87 million individuals, or 40% of the population, self-reported that they could speak Irish (CSO, 2022). While it is mandatory to learn Irish at primary and secondary levels, the reality of everyday language use outside the education system, however, tells a contrasting story. In the Republic, fewer than 72,000 people (and within the Gaeltacht - predominantly Irish-speaking - regions, only 20,261 people) use Irish daily. There are three main dialects of Irish, which are quite distinct from one another.

2.2. Learning Irish at second level

Learning Irish in English-medium schools presents challenges, with learners often feeling disengaged due to a lack of native/fluent speakers and social interaction opportunities for language acquisition. Many teachers themselves are second language learners and may lack confidence in their language proficiency. The weight of responsibility for maintaining the language can often leave them feeling unsupported and isolated (Dunne, 2015).

There is a strong emphasis on State examinations in Irish language education at second level, which often shapes the curriculum and instruction methods. The syllabus combines literary and language elements, with increasing emphasis on formal grammar tests. It covers listening comprehension, written expression, and reading comprehension. At Leaving Certificate level, the oral exam constitutes 40% of the total. Literature and written expression are also emphasised, usually assessed through essays.

2.3. Corrective feedback in iCALL

iCALL, a subfield of CALL, "applies concepts, techniques, algorithms, and technologies from artificial intelligence to CALL" (Heift, 2021). An essential challenge in iCALL is integrating teacher-student Corrective Feedback (CF) loops. CF in iCALL is often given automatically, via automatic grammar checkers and text evaluators; however, this can also be combined with manual CF, such as written/audio feedback from a teacher. It is important to find a balance among the different ways of presenting and giving CF to enhance the learning experience. Auto-generated feedback, while faster and less labour-intensive, lacks reliability (Lavolette et al., 2015), particularly in under-resourced languages; therefore, teachers need to be aware of its limitations. Various iCALL platforms have been developed to explore effective feedback methods. Ai (2017) found that using

graduated feedback by increasing specificity was the most effective method for students to self-identify and self-correct. Another iCALL platform developed by Heift (2004) concluded a combination of error highlighting and meta-linguistic information was the most effective CF process, measured by students' improvements and satisfaction. These platforms emphasize the importance of responsive feedback systems.

These platforms, among others, show that it is important for iCALL platforms to implement a feedback system that learners respond well to. *An Scéalaí* incorporates proven automatic and manual feedback features while addressing the Irish teaching community needs. The questionnaire in the Methodology section gathers teacher CF experiences, which is vital for *An Scéalaí's* future development to meet the needs of the end-user.

3. An Scéalaí

An Scéalaí is an open-source intelligent-Computer-Assisted Language Learning (iCALL) platform, which: (i) acts as a digital hub for students and teachers, integrating the latest speech technology developments; (ii) serves as a tool for researchers to collect and analyse learner language data; and (iii) informs development priorities for core TTS and ASR development for Irish. Given that Irish is an endangered language, this data is invaluable for the future development of Irish language technologies.

The *An Scéalaí* platform utilises speech and NLP tools to provide students with different resources to aid their acquisition of writing, listening, reading, and speaking. One of the core features of the platform is for students to practise their composition skills, and teachers can monitor and give CF on their written work. Students compose their 'stories', via typing or speaking (ASR), and use different tools, such as grammar checkers (*An Gramadóir*[3] and a selection of in-house checkers), TTS synthesisers[2], and dictionaries[4], to enhance their composition and self-correction. Teachers have oversight of their students' stories and can provide CF by leaving textual and/or audio comments on their student's stories or by highlighting any errors.

Once a teacher has left feedback on a student's story, the student can open their story and view the changes. The feedback and comments are displayed next to the story so the students can refer to them while correcting any errors (Fig. 1).

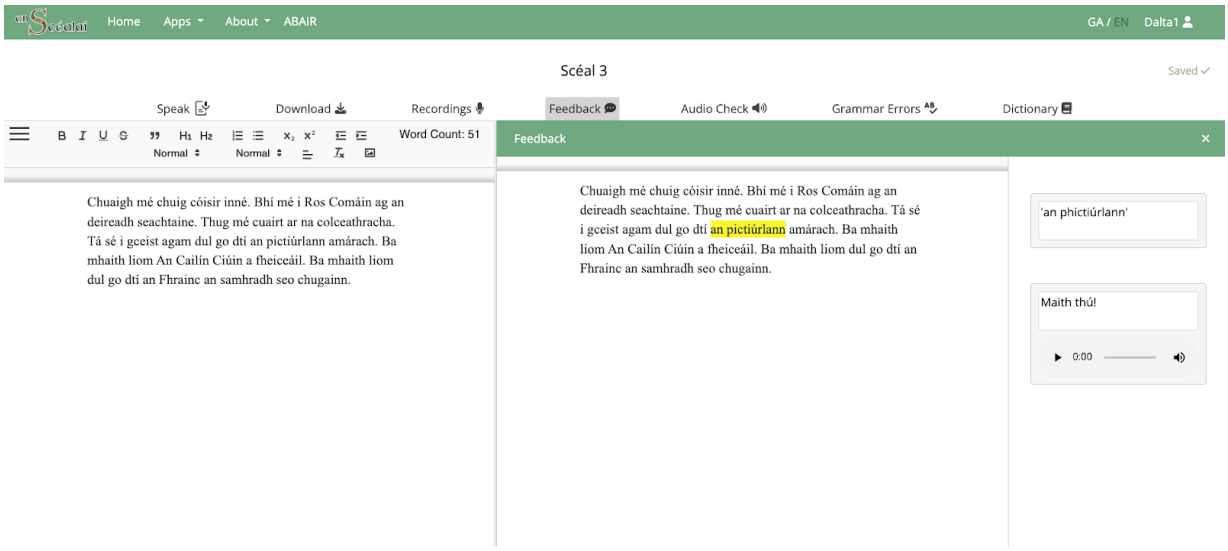


Figure 1. Students can view audio/textual feedback on their written work left by their teacher.

[3] https://cadhan.com/gramadoir/index-en.html

[4] https://www.teanglann.ie; https://www.focloir.ie; http://www.potafocal.com

In addition to viewing their students' stories, teachers can also view a statistics dashboard that shows different metrics on their performance (Fig 2). The dashboard includes graphs displaying the vocabulary variety, average story word counts, dictionary word lookups, and errors that the grammar checkers have automatically detected. These tools give teachers an overview of error trends in both individual students' performance as well as at the classroom level.

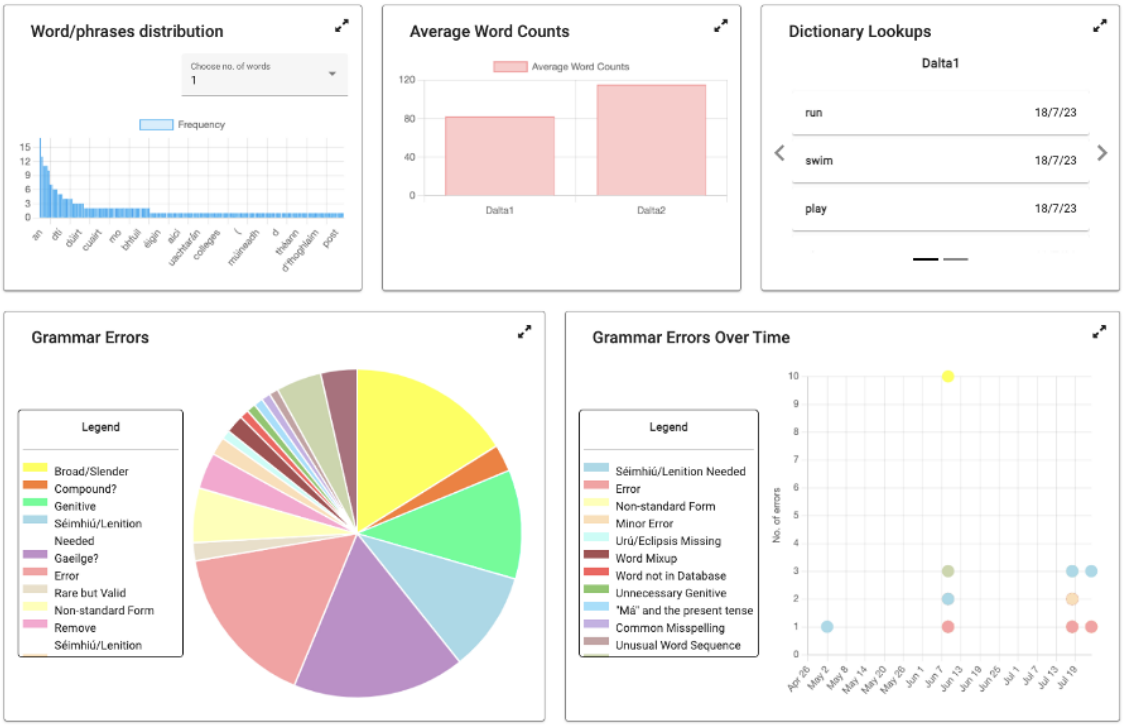


Figure 2. The teacher statistics dashboard. Available at: scealai.abair.ie

4. Methodology

A two-part (18 question) questionnaire[5] was designed and created within Google Forms and presented in a bilingual format. The questionnaire was aimed at eliciting detailed responses from second-level Irish language teachers. The first section sought general feedback on their experiences in teaching Irish, and the second section was more specifically focused on *An Scéalaí*. In this latter section, teachers were asked to watch a four-minute demo video that presented *An Scéalaí* from a teacher's perspective, and they were then asked to share their views on various features of the platform, particularly the newest addition, the statistics dashboard. The questionnaire was disseminated through a private Facebook group titled 'Múinteoirí na Gaeilge' (Teachers of Irish), which has over 8,000 members. This group of second-level teachers was chosen as it includes many active members.

5. Results and discussion

Thirty-two open-ended responses were gathered from second-level teachers of Irish (teaching Junior and Leaving certificate syllabi). Included here are some of the most notable questions and responses from the survey that can be used to direct further technical developments of *An Scéalaí*.

The questionnaire reveals teachers' strong inclination toward technology use in Irish classrooms; all but one in the sample group stated they use technology (Fig. 3). (While we believe teachers are incorporating technology into their lessons more extensively post-Covid, it is worth noting this data may have a bias towards technologically engaged teachers due to recruitment through social media during a school holiday period.)

[5] Teacher Questionnaire: https://forms.gle/h7H1ajKwspC8zmyS9

Teachers reported using numerous Irish language resources, many of which are not specifically designed for educational or language learning purposes (e.g. Irish language media).

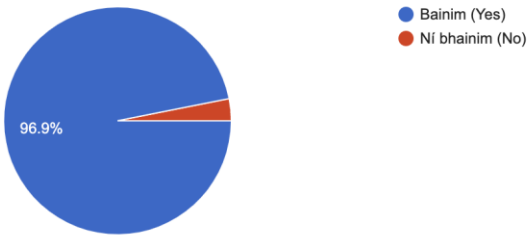


Figure 3. Teachers' (N=32) responses to: *Do you make use of any technologies when teaching Irish?*

Teachers also use numerous language-independent resources, such as Kahoot, Quizlet, and Blooket. While these resources can be useful, they require significant individual effort to tailor them for Irish. Therefore, teachers expressed strong desire for a unified platform that would supply Irish-specific materials and combine existing relevant resources in a pedagogically sound way, consolidating existing resources and reducing workload.

It is clear from the questionnaire results that teachers' approaches to platform design and development is influenced by the structure of examinations. Many respondents noted a significant disadvantage of an online platform is the necessity for students to physically write for their exams, making an extensive online engagement potentially counterproductive.

Another key area is teacher training to effectively address technological needs. Two teachers mentioned a disadvantage to the platform is the need for technological training, which mirrors that mentioned in Hubbard (2023). We can address this by creating short demonstration/instructional videos to show various features and to explain their relevance in terms of the teaching/learning of Irish.

Several teachers recognised the value of oral/aural feedback. When asked about current CF practices, over half of the teachers reported marking learner copybooks with highlights and notes and leaving audio messages for digital assignments. Six teachers mentioned using OneNote to leave CF in the form of voice messages. This is encouraging, as it aligns with what is already implemented in *An Scéalaí* (see Figure 1) and the core principle of placing the spoken language centre stage. For automatic feedback, teachers emphasised the importance of tools that are as accurate as possible, so students are not provided with misleading information.

A third of the teachers reported positive encouragement is an important part of the CF process. Currently *An Scéalaí* is dependent on manual/teacher-generated CF to provide this positive feedback for learners. In the future we would like to incorporate more positive reinforcement into our automatic feedback systems.

6. Conclusions

This paper focuses on the experiences of second-level teachers in (i) using technology for teaching the Irish language and (ii) reacting to the *An Scéalaí* platform with a view to corroborating the development trajectory of the platform.

The teacher-questionnaire feedback emphasises the importance of continually augmenting the resources available on *An Scéalaí* to meet a variety of needs. It highlights a significant demand for the development of a more diverse and engaging collection of aural and written language content as a top priority. Additionally, there is evidence that our developments align with the needs of teachers in terms of CF, namely composing individualised oral and written notes.

The engagement and buy-in of the wider teaching community are fundamental for the success of the platform. The teacher community must be 'in-the-loop' in the development of the core platform, providing input that directs development pathways, prioritising necessary applications, and actively participating in the creation and provision of these resources.

In contrast to major world languages, endangered languages require a distinctive approach. Intensive community engagement and participation become vitally important in the pursuit of preserving and revitalising languages that are under-resourced. This need becomes even more apparent when the learner's attitudes and motivation towards acquiring the language are either negative or minimal. Such a situation necessitates creating an encouraging and supportive learning environment, something that active community involvement can greatly assist in achieving.

Acknowledgements

This work is supported by An Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meáin with funding from the National Lottery as part of An Stráitéis 20 Bliain don Ghaeilge, 2010 – 2030, and An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta (COGG).

References

Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, *29*(3), 313-334.

CSO, Census of population 2022 - summary results. CSO. (2022). https://www.cso.ie/en/baile/eisiuinti/epi/p-cpr/daonaireamh2022-torthaiachoimre/tablai-anghaeilgeagusanghaeltacht/

Dunne, C. M. (2015). Becoming a Teacher of Irish: The Evolution of Beliefs, Attitudes and Role of Perceptions (Doctoral dissertation, Trinity College Dublin).

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, *16*(2), 416-431.

Heift, T. (2021). Intelligent Computer Assisted Language Learning. In Mohebbi, H. & Coombe, C. (Eds.), Research Questions in Language Education and Applied Linguistics. Springer

Hubbard, P. (2023). Emerging technologies and language learning: mining the past to transform the future. Journal of China Computer Assisted Language Learning. https://doi.org/10.1515/jccall-2023-0003

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language, Learning & Technology*, *19*(2).

Moseley, C. (2010), Atlas of the World's Languages in Danger. UNESCO.

Ní Chiaráin, N., Comtois, M., Nolan, O., Gunning, N. R., Sloan, J., Berthelsen, H., & Chasaide, A. N. (2022, June). Celtic CALL: strengthening the vital role of education for language transmission. In Proceedings of the 4th Celtic Language Technology Workshop within LREC2022 (pp. 71-76).

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16960

Listening tasks in virtual reality: A pilot study

Regina Kaplan-Rakowski[a] and Yongluan Ye[b]

[a]Department of Learning Technologies, University of North Texas, Regina.Kaplan-Rakowski@unt.edu and [b]Department of Learning Technologies, University of North Texas, YongluanYe@my.unt.edu

How to cite: Kaplan-Rakowski, R.; Ye, Y. (2023). Listening tasks in virtual reality: A pilot study. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16960

Abstract

Listening comprehension is a crucial skill in Foreign Language (FL) learning, yet many FL learners encounter challenges in this area, leading to frustration and discouragement. To address the issue, this pilot study investigated whether high-immersion Virtual Reality (VR) could effectively facilitate attentive listening tasks for FL learners. This quasi-experimental, within-subject pilot study involved ten learners of Chinese from an underserved, urban high school in the United States. Participants experienced an animated story in VR with and without captions, then completed pre- and post-tests to assess language gains and listening comprehension. A questionnaire measured participants' perceptions of VR and the use of captions. Results indicated that VR positively impacted language gains and listening comprehension, showing improvement with both groups of participants. Although the group without captions slightly outperformed the group with captions, this difference was not statistically significant. Novice learners reported challenges in comprehending the VR story, highlighting their need for appropriate scaffolding. Overall, the findings suggest that VR can be an effective tool for practicing listening comprehension for FL learners, but careful consideration of content complexity and scaffolding is crucial for learners at different proficiency levels. This study lays the groundwork for further research on VR-assisted language learning.

Keywords: *high-immersion Virtual Reality (VR), Virtual Reality–Assisted Language Learning (VRALL), listening comprehension, captions, language learning.*

1. Introduction

The development of listening skills is essential in Foreign Language (FL) learning, but many FL learners find listening comprehension challenging (Feyten, 1991; Graham, 2006). To succeed in listening comprehension tasks, FL learners need to identify individual words while forming mental representations. When failing to do so, learners often neglect the rest of the message, thus aggravating the accumulation of difficulties (Goh, 2000) and making the learners feel discouraged.

Given that learners find high-immersion Virtual Reality (VR) enjoyable and motivating (Kaplan-Rakowski & Gruber, 2022, 2023), the idea of providing listening comprehension tasks in VR to encourage learners to listen attentively is compelling although understudied (Dhimolea et al., 2022). Studies on listening skills attempted to use VR to prompt information retention (Lee, 2019; Pinto, 2019; Tai & Chen, 2021). Tai and Chen (2021) found that learners who experienced the interaction in VR retained significantly more knowledge than those who watched two-dimensional videos of the same interactions. Similar to Tai and Chen (2021), Pinto (2019) reported that learners had a higher sense of presence and engagement in VR, but VR and two-dimension (2D) video approach

made no significant difference in learners' knowledge retention. Meanwhile, the small sample size (N = 12) and the short language input (~2 minutes) limited the generalizability of Pinto's (2019) findings.

Although previous studies on the effect of VR on listening comprehension yielded promising results (Lee, 2019; Pinto, 2019; Tai & Chen, 2021), they mainly focused on understanding dialogues instead of monologues. Monologues are essential for FL learners to function in the target language environment. However, monologues provide fewer cues that can aid comprehension of the materials (Fox Tree, 1999). Only Lee (2019) examined the effect of VR on monologues without incorporating the interactive functions of VR, limiting the impact of the technology.

The use of captions has shown significantly positive results in helping FL learners improve their listening comprehension (Ghasemboland & Nafissi, 2012; Hsu et al., 2013), but the impact of captions in VR is still uncertain. While VR provides more visual details to the learners, Mayer and Pilegard (2014) argued that too much information delivered through the visual channel may cause essential overload and impede learning outcomes. Adding captions to the rich visual information in VR may increase the cognitive load and hinder FL learners' listening comprehension.

Study Purpose and Research Questions

To fill the gaps in the literature, this pilot study aimed to prepare for a more in-depth project exploring the effect of VR on listening comprehension and the effect of captions in VR (see Ye & Kaplan-Rakowski, 2023). The study seeks to answer the following questions: (1) Does VR promote language gains and listening comprehension? (2) How do captions affect listening comprehension in VR? and (3) Will captions cause essential overload or support listening in VR?

2. Method

This pilot study was conducted in preparation for a larger project investigating the impact of VR on the development of listening comprehension skills (Ye & Kaplan-Rakowski, 2023). To answer the research questions, we designed a quasi-experimental, within-subject study. Parametric data were analyzed using independent *t*-tests, and nonparametric data were analyzed using descriptive statistics.

2.1 Participants

The study participants (N = 10) came from low-income families residing in an urban area of the United States. Eleven participants consented to participate, but only ten (male = 6, female = 4) completed all parts of the study. All participants were 18 years old but varied greatly in their years of Chinese language education. The majority of participants had been learning Chinese for four to six years, with the least experienced participant having studied Chinese for one year and the most experienced having studied it for over ten years. This variety was to help gain data on the learning experience in VR from learners of different language proficiency.

2.2. Procedure

The procedure used in the study had three stages. First, the participants received scaffolding (see Figure 1) on the story used in the experiment including explanations of the six high-frequency vocabulary items from the story and pictures showing the setting and the main characters. Afterward, participants took a pre-test to establish their language knowledge baseline.

Figure 1. Teacher providing scaffolding with sceenshots from *The Line*. Photo by the authors.

Second, the participants experienced the target story in VR (see Figure 2) in two parts. This was an animated, romantic story which was embedded within the app called *The Line*. We chose this particular app because the story it offers has a repetitive structure that may help comprehension. Moreover, the interaction is simple and unlikely to cause motion sickness. The video version of this story can be viewed here. The participants were randomly assigned to the caption or non-caption conditions to experience the first part of the story and then switch to the other condition to experience the second part of the story. See Figure 3 for a screenshot of the story.

Figure 2. Participant experiencing the virtual reality story *The Line*. Photo by the authors.

Figure 3. A screenshot of the interactive virtual reality story with captions

Third, after each intervention, the participants took a post-test to determine their language gains. Finally, the participants completed a questionnaire that collected demographic data and data that reflected the participants' experience in VR.

2.3. Instruments

The present study used four instruments: a pre-test, two post-tests, and a questionnaire. The pre-test and post-tests contained 14 multiple-choice and three translation questions, in which participants needed to listen to audio recordings to determine the meanings of words, phrases, and sentences. The post-tests also had three inferential questions regarding the story that could not be answered by merely watching the VR story. The participants' answers were scored and the accuracy rate was used in the analysis. The pre- and post-tests assessed only participants' listening comprehension, meaning that participants did not need to read Chinese in the tests.

The questionnaire collected participants' demographic data such as gender, age, and years of learning Chinese. To reflect the participants' experience using VR, the questionnaire used questions with a seven-point Likert scale. The questionnaire adapted four questions from the instrument developed by Leppink et al. (2014) to measure the cognitive load, three questions from the instrument by Ellis et al. (2021) to measure the positive effect of VR, and four questions from the instrument by Gandolfi et al. (2021) to measure presence.

3. Results and Discussion

Each test question underwent validation using Cronbach Alpha, which is a method to measure the reliability of questionnaires. It assesses whether scores from multiple items can reliably capture the target information. All pre-test questions were valid, but the post-test achieved only 78.6% validity because of three open-ended inferential questions. During the follow-up interview, several participants reported that those questions were too difficult because they asked for details that were mentioned in the narration only once. Therefore, before conducting the larger and main study (Ye & Kaplan-Rakowski, 2023), inferential questions were changed to multiple choice questions asking for opinions and feelings of the characters from the story.

The free-response portion of the post-tests was scored by two researchers following the same rubric: each correctly translated key item earned one point. Missed items or wrong translations earned zero points. When disputes occurred, the two researchers discussed the issue at hand until an agreement was achieved. The accuracy rate of each participant was analyzed using a *T*-test. The participants who experienced the story with captions on average

increased by 6.1 points on the listening comprehension test (*SD* = 9.96), and those who experienced the VR story without captions increased by 8.1 points (*SD* = 12.16). The non-caption group insignificantly outperformed the caption group [$t(10) = 0.581$, $p = 0.078$]. In addition, two participants' scores were lower in one or both post-tests, which suggested that these learners guessed responses in the pre-test although the researchers urged them to choose the "I don't know" option when needed.

Nonparametric data were analyzed and the descriptive statistics (means, standard deviations, theoretical minimum, and maximum range of scores for each item) are presented in Table 1. Two participants at the novice level in Chinese reported greater difficulties comprehending the test questions and the VR story than the others. The 11th participant, who did not finish the study, decided to leave because the test was too intimidating. Therefore, the researchers decided to exclude novice Chinese learners from the main study.

Table 1. Descriptive statistics of the study constructs

N = 10	M	SD	Min	Max
Listening comprehension-VR with captions	6.10	9.96	-13.00	21.00
Listening comprehension-VR without captions	8.10	12.16	-16.00	26.00
Caption usefulness	5.28	1.21	4.00	6.25
VR positive effect	6.25	0.20	6.00	6.50
Cognitive load	4.33	0.57	3.33	4.67
Sense of presence	5.63	0.92	4.00	7.00

Note. Score ranges were 0 to 100 for listening comprehension tests, 1 to 7 for caption usefulness, 1 to 7 for positive affect, 1 to 7 for cognitive load, and 1 to 7 for sense of presence. Some theoretical minimal values are below zero. This is because some students did worse on the post-test than on the pre-test.

The results indicate that VR can support listening comprehension amongst Chinese FL learners, which answers our first research question. This learning outcome may be explained by the participants' enjoyment in using VR as a learning media and that VR created a reality in which participants felt present. To answer the second question, although a majority of the participants considered captions helpful in aiding comprehension, the learning outcomes of the caption group were insignificantly lower than of the non-caption group. This result may have been caused by some participants' low Chinese proficiency level and their limited ability to process the captions at the pace of the VR story. This study did not measure reading comprehension that may be enhanced with captions.

This pilot study was based on a small sample size. Future studies should reexamine the claims with more participants. Moreover, because of the cost of VR devices, VR may need to outperform other lower-cost media (e.g. computers, tables) in supporting listening comprehension for the technology to be deemed desirable to learners and educators. Therefore, future research should examine the impact of alternative media on listening comprehension.

4. Conclusions

This study explored the effect of VR on listening comprehension and validated the methodology used in the main study (Ye & Kaplan-Rakowski, 2023). Overall, participants reacted positively toward learning language with VR, and the learning outcome suggested that the VR experience supported their learning. Learning outcomes with different media will be examined in the main study.

While all participants reported a positive effect toward VR, those at the novice level in Chinese found the tests too difficult. One participant complained about the large quantity of unfamiliar vocabulary and the scaffolding, which, although helpful, was insufficient. Given previous research on learning vocabulary in VR (Papin & Kaplan-

Rakowski, 2022), vocabulary and grammar may need to be limited in input materials and assessments for novice language learners.

Acknowledgements

We extend our thanks to all the study participants for their willingness to participate, share insights, and dedicate their time to advance our research.

References

Christmann, A., & Van Aelst, S. (2006). Robust estimation of Cronbach's alpha. *Journal of Multivariate Analysis, 97*(7), 1660–1674.

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of virtual reality language learning. *TechTrends, 66,* 810–824. https://doi.org/10.1007/s11528-022-00717-w

Ellis, M. E., Downey, J. P., Chen, A. N., & Lu, H. (2021). Why Taiwanese seniors use technology. *Asia Pacific Management Review, 26*(3), 149–159. https://doi.org/10.1016/j.apmrv.2021.01.001

Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal*, *75*(2), 173–180. https://doi.org/10.2307/328825

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*(1), 35–53.

Gandolfi, E., Kosko, K. W., & Ferdig, R. E. (2021). Situating presence within extended reality for teacher training: Validation of the extended Reality Presence Scale (XRPS) in preservice teacher use of immersive 360 video. *British Journal of Educational Technology*, *52*(2), 824–841.

Ghasemboland, F., & Nafissi, Z. (2012). The effects of using English captions on Iranian EFL students' listening comprehension. *Procedia, Social, and Behavioral Sciences, 64,* 105–112. https://doi.org/10.1016/j.sbspro.2012.11.013

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*(2), 165–182.

Hsu, C., Hwang, G., Chang, Y., & Chang, C. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society, 16*(1), 403–414.

Kaplan-Rakowski, R., & Gruber, A. (2023). An experimental study on reading in high-immersion virtual reality. *British Journal of Educational Technology, 00,* 1–19. https://doi.org/10.1111/bjet.13392

Kaplan-Rakowski, R., & Gruber, A. (2022). Motivation and reading in high-immersion virtual reality. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 208–213). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1460

Lee, A. (2019). Using virtual reality to test academic listening proficiency. *Korean Journal of English Language and Linguistics, 19*(4), 688–712.

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30,* 32–42.

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). Cambridge University Press.

Papin, K., & Kaplan-Rakowski, R. (2022). A study on vocabulary learning using immersive 360° pictures. *Computer Assisted Language Learning, 35*. https://doi.org/10.1080/09588221.2022.2068613

Pinto, D., Peixoto, B., Krassmann, A., Melo, M., Cabral, L., & Bessa, M. (2019, April). Virtual reality in education: Learning a foreign language. In Á. Rocha, H. Adeli, L. Reis, & S. Costanzo (Eds.), *World Conference on Information Systems and Technologies* (pp. 589–597). Springer.

Tai, T.-Y., & Chen, H. H.-J. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research, 59*(7), 1272–1293. https://doi.org/10.1177/0735633121994291

Ye, Y., & Kaplan-Rakowski, R. (2023). Practicing listening comprehension skills in high-immersion virtual reality. *SSRN*, https://ssrn.com/abstract=4335690

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16914

Machine translation as a potential tool for comprehension in multilingual classes

Marni Manegre[a], Mar Gutiérrez-Colón[b] and Andreu Martí Aguilar[c]

[a]Department of English and German Studies, Universitat Rovira i Virgili, marnilynne.manegre@urv.cat; [b]Department of English and German Studies, Universitat Rovira i Virgili, mar.gutierrezcolon@urv.cat and [c]LMS Manager and Instructional Design, Universitat Rovira i Virgili, andreu.marti@urv.cat

How to cite: Manegre, M., Gutiérrez-Colón, M.; Martí Aguilar, A. (2023). Machine translation as a tool to improve comprehension in a university lecture. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16914

Abstract

With an increase in international student enrollment and participation in student mobility programs year over year, the goal of this project was to determine whether Microsoft Translate could be used as a simultaneous translation tool in the classroom to help foreign students understand course material, especially in situations where the instructor is not lecturing in a language the students are fluent in. This was tested in two master's level courses at a Catalan institution with mostly international students. The results suggest that the language of instruction, which was either English or Catalan, determined whether the participants perceived to comprehend the translation. Furthermore, speech style was essential, with scripted speech appearing to outperform unscripted speech. The goal of this study was to detect whether this program might be utilized to make university lectures more accessible to all students. The findings suggest that when the teacher uses scripted speech in the classroom, students of various language backgrounds seem to understand the key ideas of the lectures.

Keywords: *speech-to-text translation, machine translation, Microsoft Translate, automatic speech recognition.*

1. Introduction

Human translators are trained to interpret complicated texts and consider aspects such as tone and culture or other nuances of a language; however, Machine Translation (MT) can often be free or inexpensive. Due to the fiscal cost of human translators, businesses have identified "a critical need to develop simultaneous MT techniques to reduce the burden of human interpreters and to make simultaneous interpreting services more accessible and affordable" (Wang et al., 2021). Machines are adept at analyzing massive volumes of data and may quickly deliver translations for simple, easy texts for a fraction of the cost.

The basic goal of MT is to break down linguistic barriers (Vieira et al., 2021), and this can be done with the help of two classes of translation. When the user must enter the text before the translation shows, this is known as *consecutive translation*. In contrast, *simultaneous translation* happens as the speaker is speaking (Cho & Esipova, 2017). Simultaneous MT has the advantage of allowing speakers from various language origins to engage in the same meetings or lectures, notwithstanding their differences in linguistic backgrounds. However, there are several difficulties to solve while using MT (Zakir & Nagoor, 2017). It might be challenging to

interpret human speech during simultaneous translations due to numerous pauses and interruptions. The translation tool would need to remove redundancies and repetitions before adding the punctuation in order for the sentences to be read and fully understood. The instrument must be congruent with the manner of speech, whether official and professional or informal and casual, in addition to arranging the sentence structure. One of the major barriers to MT is the large number of human languages. Many translation algorithms are incapable of detecting languages spoken in distant areas by few native speakers.

1.1. Using simultaneous machine translation in university lectures

There are various reasons why MT should be considered as simultaneous translation technology in university courses. For students who do not speak the language of instruction, MT can be utilized to give real-time translation of lectures. This is especially effective in classes with a high proportion of international students (Yang et al., 2021; Muller et al., 2016; Cho et al., 2013; Wolfel et al., 2008).

Additionally, MT may be utilized to aid with language learning. To assist students who are still learning the language of instruction, instructors may provide the students with translated versions of their lectures. This can aid in their comprehension of the material as well as their language abilities (Paterson, 2022).

Finally, MT enables accessibility for all pupils. MT can also be used to make lectures more accessible to deaf or hard-of-hearing students. These students can better follow along with the lecture if real-time subtitles or translations are provided (Kawas et al., 2016).

1.2. Introduction to the Problem

Catalan, a language spoken by around eight million people, is the official language of Catalonia's universities. Catalonia's university system, which consists of 12 universities, has an exceptional capacity to recruit international students. This is evident as there are now 34,755 international students among the 289,369 students registered in Catalan institutions. These students represent 176 different countries. Additionally, Catalonia has 150 undergraduate and master's degree programs that are exclusively taught in English (Recerca Universitats, 2022).

It sometimes occurs that these students from abroad enroll in subjects that are not taught in English, either because they need additional credits or because they are enrolled in Spanish or Catalan courses and assume they can attend a subject in those languages. Moreover, certain university programs provide lessons in all three languages (English, Catalan, and Spanish), which may not be accessible to all students if there are linguistic barriers. The students may even lack a good level of English or one of the other languages, making it difficult for them to grasp the lecture's topic.

The key difficulty is determining how to make the classes available to all enrolled students. In 2022, the Servei de Recursos Educatius of the Catalan universities was asked by the Catalan Government (via the CSUC-Consortium of Catalan Universities) to test MS Translate in classes containing a high number of international students. The objective was to see whether this simultaneous translation tool could solve this linguistic problem.

As a result, the main research question is as follows: Can Microsoft Translate be used in multinational classrooms to break down linguistic barriers? More specifically, does the style of speech, either unscripted or scripted, influence whether students feel they understand the lecture? Additionally, does the language of instruction appear to function better with the software, either English or Catalan?

2. Method

2.1. Participants

The participants were students in two master’s courses at a university in the Catalan region of Spain. There were 29 participants in total, with 14 enrolled in a degree in international markets and 15 in the degree of teaching English as a foreign language. The students reported their native languages to be English, Catalan, Arabic,

Riffian, Ukrainian, Filipino, Russian, Polish, Spanish, Portuguese, Chinese, Romanian, Italian, German, Persian, Slovak, and Bahasa. Additionally, all of the students have a B2 level of English or higher. Only four of the students were proficient in Catalan.

2.2. Materials

The participants were instructed to use their own devices. They were told to log into Microsoft Translate with the class code and select their native language. The instructors wore lapel microphones that were also connected to the program. A technical advisor was available in the class to help the students log in and to make sure the application was functioning properly.

2.3. Procedure

The instructor, who was a native speaker of English, lectured in English, while the other instructor, who was a native speaker of Catalan, lectured in Catalan. The English-speaking instructor lectured first and read a news story in English. The same story was read by the Catalan instructor immediately following. Then, each of the instructors discussed the topic of mobile-assisted language learning in an unscripted fashion. The students responded to a Likert questionnaire scoring the speech based on whether they perceived the translation was easy to follow, whether the tool was easy to use, and whether they would recommend using this tool in a university classroom. The Likert scale was a seven point scale where one was considered to be strong disagreement with the statement and seven to be strong agreement with the statement. The data from the Likert scale was converted to a csv file and analyzed using JASP statistical software.

3. Results

Since there were two classes from different departments participating in this study, an ANOVA was conducted to see whether the two classes were responding in a similar fashion. There was no significant difference between the two classes of students $F(1) = 0.827$, $p = 0.371$; therefore, both classes were analysed together.

As all of the participants went through the same lectures in the procedure, a repeated measures ANOVA was used to compare the results for the different languages (English and Catalan) and the different styles of speech (scripted and unscripted). The results show that the language of instruction was significant with a medium to high effect $F(1) = 23.417$, $p < .001$, $\eta^2 = 0.074$, where the students scored the translations of English to be higher than the translations of Catalan. Additionally, there was a significant difference between the styles of speech with a low to medium effect $F(1) = 10.487$, $p = 0.003$, $\eta^2 = 0.029$.

As previously stated, the language altered the students' overall views when using Microsoft Translate. We can observe the means for each of the students' perceptions, the style of lecture, and the language when we look at the descriptive statistics. A score of one would be the lowest, a score of seven would be the highest, and a score of four would be neutral. For English, all mean ratings were greater than four, indicating that students typically believed the translation was accurate to their native languages, that they understood the translation, that the tool was simple to use, and that they would suggest it for university courses. With the exception of the unscripted Catalan lecture, the majority of the scores for the Catalan lectures were above neutral and in agreement. Most of the students were uncertain whether the unscripted Catalan translation was accurate (see Table 1 below).

Table 1. Descriptive Statistics.

Language	Style	Perception	Mean	SD
Catalan	Free Speech	Correct	3.931	1.486
		Easy	5.724	1.360

		Recommended	4.966	1.955
		Understood	4.690	1.538
	Scripted	Correct	4.828	1.560
		Easy	5.724	1.645
		Recommended	5.276	1.888
		Understood	5.172	1.649
English	Free Speech	Correct	4.793	1.590
		Easy	6.000	1.309
		Recommended	5.414	1.803
		Understood	5.690	1.390
	Scripted	Correct	5.517	1.153
		Easy	6.034	1.322
		Recommended	5.655	1.838
		Understood	6.034	1.149

4. Discussion

Machine learning and artificial intelligence advancements have considerably increased the quality and speed of speech-to-text translation, transforming it into a powerful tool for communication, automation, and accessibility. Since there is an increase in international enrollment year after year in Catalan universities, this study was designed to assess if students thought a speech-to-text translation technology, particularly Microsoft Translate, was useful for comprehending lectures given in a foreign language.

The first research question investigated whether the manner of speech, scripted or unscripted, influenced the students' perceptions of translation accuracy. The results demonstrate that the style of speech made a difference for both languages, $F(1) = 10.487$, $p = 0.003$, with scripted speech being significantly easier to follow than unstructured speech and scripted speech in English being easier to follow than scripted speech in Catalan.

To answer the second research question, whether English appears to perform better in a synchronous translation with Microsoft Translate than Catalan, which has a much smaller number of native speakers, the results show that the language used in the application made a difference, with English translations being considered more accurate, $F(1) = 23.417$, $p < .001$.

4.1. Limitations of the Study

The main limitation of the study is that only a fraction of global languages were included. Additionally, the assessment was based on the user experience and their perceptions of whether they believed this tool would work for comprehension of a university lecture, and such perceptions may hold biases. The assessment did not consider the actual accuracy of the translations and there were no controls to determine whether the students comprehended the lecture with and without MS Translate. Future research should include more variables; such

as attention disruption (reading in one language while listening to a different one), the influence of the proficiency level of the language that is being translated, and an assessment of the accuracy of each translation from the first languages. Further, a larger sample could be included in future studies where there are more controls.

5. Conclusions

The findings indicate that the application appears to be appropriate for use in university courses. The spontaneous translations of certain sentences caused some difficulty among the students, but they typically comprehended the lectures. The results and feedback from this study indicate that MS Translate as a simultaneous translation tool performs well in a classroom setting for both scripted and unscripted English translations. MS Translate has been demonstrated to perform well for scripted speech in Catalan, and it appears to work for unstructured speech as well; however, accuracy cannot be determined from this study. While the quality of MT tools and machine learning has considerable room for improvement, particularly in terms of accuracy, speed, and context awareness, speech-to-text machine translations will continue to become more accurate and coherent as machine learning develops.

Acknowledgements

We would like to to acknowledge the Universitat Rovira i Virgili's Servei de Recursos Educatius, that gave us the opportunity to conduct the study.

References

Cho, K., & Esipova, M. (2016). Can neural machine translation do simultaneous translation?. *arXiv preprint arXiv:1606.02012*.

Cho, E., Fugen, Ch., Hermann, T, Kilgour, K., Mediani, M., Mohr, Ch., Niehues, J., Rottmann, K., Saam, Ch., Stuker, S., & Waibel, S. (2013). A Real-World System for Simultaneous Translation of German Lectures. INTERSPEECH 2013. International Speech Communication Association. https://www.isca-speech.org/archive_v0/archive_papers/interspeech_2013/i13_3473.pdf

Kawas, S., Karalis, G., Wen, T., & Ladner, R.E. (2016). Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. ASSETS '16: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility October 2016. 15–23. https://doi.org/10.1145/2982142.2982164

Muller, M., Nguyen, T. S., Niehues, J., Cho, E., Krüger, B., Ha, Th., Kilgour, K., Sperber, M., Mediani, M., Stüker, S., & Waibel, A. (2016). Speech translation framework for simultaneous lecture translation. Proceedings of NAACL-HLT 2016, pages 82–86, San Diego, California. Association for Computational Linguistics. https://aclanthology.org/N16-3017.pdf

Paterson, K. (2022). Machine translation in higher education: Perceptions, policy, and pedagogy. *TESOL Journal*, e690. https://doi.org/10.1002/tesj.690

Recerca Universitats (2022), Web page: https://govern.cat/salapremsa/notes-premsa/421883/sistema-universitari-catala-estrena-nou-web-internacional-coincidint-fira-americana-deducacio-superior-que-se-celebra-denver

Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, *24*(11), 1515-1532. https://doi.org/10.1080/1369118X.2020.1776370

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2021). Progress in Machine Translation. *Engineering.* https://doi.org/10.1016/j.eng.2021.03.023

Wolfel, M., Kolss, M., Kraft, F., Niehues, J., Paulik, M., Waibel, A.; (2008). Simultaneous machine translation of german lectures into english: Investigating research challenges for the future. 2008 IEEE Spoken Language Technology Workshop, Goa, India. 233-236. https://doi.org/10.1109/SLT.2008.4777883

Yang, Y; Wang, X; Yuan, Q. (2021). Measuring the usability of machine translation in the classroom context. Translation and Interpreting Studies. *The Journal of the American Translation and Interpreting Studies Association*, *16* (1). 101 – 123. https://doi.org/10.1075/tis.18047.yan

Zakir, H. M., & Nagoor, M. S. (2017). A brief study of challenges in machine translation. *International Journal of Computer Science Issues (IJCSI)*, *14*(2), 54. https://doi.org/10.20943/01201702.5457

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.17007

Evaluating the effectiveness of Microsoft Transcribe for automating the assessment of pronunciation in language proficiency tests

Carey Nelson[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, Montreal, , nelson.carey@uqam.ca and [b]Department of Education, Concordia University, Montreal, , walcir.cardoso@concordia.ca

How to cite: Nelson, C.; Cardoso, W. (2023). Evaluating the effectiveness of Microsoft Transcribe for automating the assessment of pronunciation in language proficiency tests. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.17007

Abstract

Improvements in Automatic Speech Recognition (ASR) have created opportunities for using it as a tool to facilitate second and foreign language (L2) assessment. These technical improvements have not only enabled automation of language proficiency test scoring but also reduced evaluator bias and errors, decreased processing time, and lowered costs for testing organizations. The purpose of this study was to evaluate English as a Second Language (ESL) pronunciation using the ASR feature in the Microsoft 365 product suite, Transcribe (MS-T). The study involved adult ESL learners at a Canadian university that partook in a language proficiency test. We examined the audio recordings of 56 candidates during the pronunciation portion of the test. Building on previous studies that found a strong correlation between scores from Google Voice Typing and human raters, the current study conducted a similar analysis comparing scores derived from MS-T to both human ratings and Google Voice Typing. Our findings indicate that the ASR capabilities of MS-T, similar to Google Voice Typing, can assume an important role in L2 speaking assessment by providing objectivity and reliability to the testing process, expediting scoring, and reducing costs.

Keywords: *automated evaluation, Automatic Speech Recognition (ASR), Language assessment, ESL pronunciation evaluation, Microsoft Transcribe (MS-T), placement tests.*

1. Introduction

Language proficiency testing is an important field that tries to meet the demand of evaluating candidates' proficiency in L2 for purposes such as employment, immigration, and academic admissions. These tests may be internationally recognized, e.g. TOEIC, IELTS, Duolingo, Versant, or created locally by individual institutions. In either case, language proficiency testing often touches on the assessment of a combination of language skills to determine the test-takers' abilities in that language. However, assessments with subjective components that require human raters to manually apply scoring rubrics to students' oral performances can be extremely time-consuming and labor-intensive (Coombe et al., 2020). Moreover, relying on human raters to manually score assessments leaves room for potential errors and inconsistencies based on subjective interpretations of scoring rubrics (Inbar-Lourie, 2017). There are also time and cost considerations associated with having human raters evaluate students' oral performance.

An ASR system can be particularly useful for language proficiency tests where multiple evaluators may be involved (Bernstein et al., 2010). Firstly, traditional language proficiency tests, e.g. TOEFL and IELTS, have

limitations in terms of providing an accurate assessment of a learner's language proficiency. These tests rely on human evaluators who may be subject to biases and inconsistencies, and who may not be able to provide an objective and standardized evaluation of language proficiency (Xerri & Briffa, 2018). In contrast, ASR technology has the potential to provide a less biased, more consistend and standardized assessment of language proficiency, which can reduce variability in scores across different evaluators (Mroz, 2020). Secondly, due to significant technological advancements in recent years, ASR has been used for various applications, including speech recognition in smartphones, virtual assistants, and language learning platforms (McCrocklin & Edalatishams, 2020). Tools that use ASR dictation, such as MS-T, offer a promising avenue for a cost-effective solution. The possibility of implementing ASR technology could help post-secondary institutions operate efficient and effective proficiency tests.

Considering a recent study by Johnson et al. (2022), who found a strong correlation between scores assessed by Google Voice Typing (GVT) and human raters, demonstrated that GVT provided reliable and valid results in evaluating participants' oral performance on a set of phonological criteria (e.g. phonemic accuracy, stress) and overall proficiency (e.g. comprehensibility) – on par with human raters. This raises the question: Can the same conclusions be drawn with other ASR tools? In a study from 2017, Këpuska and Bohouta found that GVT offered a superior recognition to Microsoft's ASR. Given the rapid advancements in speech technology over recent years, does this comparison still hold true today? Our study thus aimed at building on these previous findings by examining whether a similar ASR application, e.g. Microsoft Transcribe (MS-T; found in the Microsoft 365 product suite), has the potential to improve the validity efficiency of L2 speaking assessment. While MS-T is not as accessible to the general public as GVT, this study utilized MS-T to offer an alternative to universities that have access to Microsoft products. In addition, the choice to use MS-T will allow us to gain additional insights into the impacts of ASR on assessment through the analysis of an alternative platform. Finally, this decision will allow us to compare the performance of two predominant ASR platforms, MS-T and GVT. As such, our study aimed to answer the following Research Questions (RQs):

1. What is the relationship between MS-T-rated pronunciation scores and human-rated pronunciation scores (RQ1)?
    a. Do relationships vary between MS-T-rated scores and human-rated scores across the set of evaluation criteria used by the human raters?
    b. Do relationships vary between MS-T-rated scores and human-rated scores across participant proficiency levels?
2. How does MS-T fare compared to GVT at pronunciation scoring across the set of evaluation criteria used by the human raters (RQ2)?

2. Method

2.1. Context and participants

The sample of participants comprised 56 adults (n = 56; 21 males, 35 females; mean age: 28.1), with the following distribution of native languages: French (n=39, 68.4%), Spanish (n=4; 7.0%), Arabic (n=3; 5.6%), and others (5; 8.92%). They were undergraduate students who had taken a proficiency test either to enroll in proficiency-appropriate ESL classes or to fulfill linguistic requirements for their academic programs. They had oral production levels ranging from A1 to C2 according to the Common European Framework of Reference (CEFR) for languages (Council of Europe, 2001). The sample for this study comprised 75 recordings drawn from over 20,000 proficiency tests administered between 2015 and 2020. Only the pronunciation portion of these tests were considered, which were randomly selected across six proficiency levels, based on the candidates' pronunciation scores obtained on the tests. Nineteen sound files were eliminated from the sample because they were not clearly audible.

2.2. Procedure

This study made use of secondary data, which were originally collected by a modern language department at a French-speaking university in Canada. It was part of a larger project to implement new rubrics for scoring pronunciation on the ESL proficiency test. With respect to the pronunciation section, the participants were given two practice sentences to be read aloud. This ensured that they understood the evaluation activity procedure and that the computer system was functioning correctly. The test then gave the participants five sentences that appeared one after the other, after 20 seconds, with increasing levels of speaking and pronunciation difficulty. The students read these five sentences where the first one was a baseline sentence (same for all students) followed by four sentences that were randomly chosen from a pool organized by proficiency level. The test saved each sentence as an individual recording to be consulted by the raters. Due to an agreement to protect the integrity of the test, the baseline question cannot be shown here. Table 1 is an example of sentences to be read aloud by students and is taken from Johnson et al. (2022). The human raters used a rubric assessing five phonological components: *Comprehensibility, Phonemes, Connected speech, Word stress and rhythm,* and *Thought groups, sentence stress, and intonation*.

Table 1. Sample sentences.

Level	Sentence
1	[Baseline sentence]
2	*A trio sings to the audience as it streams onto the busy street in the cold rain.*
3	*These are more sophisticated pictures, aimed at a particular kind of filmgoer. Is she sure that this audience understands them?*
4	*After the stems are cut off the mushrooms, they are then going to be sautéed with a small onion, a clove of garlic, and an eighth of a cup of breadcrumbs.*
5	*Even though the trailer has been cleaned, there are still lingering traces of acetones and other toxic amalgams either in the gaskets or in the valve assembly.*

2.3. Procedure

For data analysis, the MS-T score, the GVT score, the final human-rater score, as well as the scores for each criterion were entered into SPSS 29. These results help to answer the research questions of whether a relationship exists between MS-T rated pronunciation and human raters' evaluations.

3. Results

Looking at the relationship between MS-T and the human-rated scores, the results show a statistically significant strong correlation between the two variables, $r_s(54) = .79$, $p < .001$. For the first sub-question pertaining to the relationships between the MS-T scores and the rubric criterion (RQ1a), the results indicate that there are statistically significant strong correlations between the MS-T score and each of the criteria considered for the assessment of pronunciation (Tables 2 and 3 show the summary of the correlations). With respect to the question about the relationship between MS-T scores and test-taker proficiency (RQ1b), results indicate a significant correlation between the MS-T and human-rated scores for lower-proficiency test takers, $r_s(54) = .59$, $p < .006$, but a non-significant weak correlation between the MS-T and human-rated scores for higher-proficiency test takers, $r_s(54) = .29$, $p = .89$.

Table 2. Correlations between MS-T score and human-rated scores by criteria.

Rubric criteria	r_s	95% BCa Cis
Comprehensibility	.83**	.72, .90
Phonemes	.76**	.62, .85
Connected speech	.78**	.65, .87
Word stress and rhythm	.73**	.57, .84
Thought groups, sentence stress, and intonation	.76**	.62, .86

Note. Confidence intervals based on 1000 bootstrap samples. ***p* < .001.

Table 3. Correlations between MS-T score and human-rated scores by proficiency level.

Rubric criteria	r_s	95% BCa Cis
Lower-level proficiency	.59**	.19, .82
Higher-level proficiency	.29	-.56, .57

Note. Confidence intervals based on 1000 bootstrap samples. ***p* < .006.

RQ2 asked about the relationship between MS-T and GVT at pronunciation scoring. The results in Table 4 show that GVT is marginally stronger, but not significant, at the criteria of *comprehensibility*, *phonemes*, and *thought groups, sentence stress, and intonation*, whereas MS-T is a little stronger at the criterion *connected speech* and *word stress and rhythm*. After running a two-tailed t-test, we observe that the mean for MS-T is significantly higher than for GVT for pronunciation scoring total $t(54) = -3.7$, $p = .001$.

Table 4. Comparison (r_s) between MS-T scores and GVT scores.

Rubric criteria	MS-T	GV-T
Comprehensibility	.83**	85**
Phonemes	.76**	.78**
Connected speech	.78**	.72**
Word stress and rhythm	.73**	.71**
Thought groups, sentence stress, and intonation	.76**	.79**
Total	.79**	.78**

Note. Confidence intervals based on 1000 bootstrap samples. ***p* < .001.

4. Discussion and conclusions

This study evaluated the effectiveness of MS-T as an ASR for scoring L2 pronunciation assessment, and compared it to another popular ASR platform, GVT. The first research question looked at how MS-T compared to human raters when evaluating L2 pronunciation samples. Our findings for RQ1 indicate that MS-T displays strong, significant correlation across each of the different phonological criteria used in the evaluation. These results indicate, like what Johnson et al. (2022) found with GVT, that there exists a strong association when evaluating L2 pronunciation between MS-T and human-rated scores. This leads us to believe that the ASR MS-T gives similar results for evaluating L2 pronunciation samples.

With regards to RQ2, we looked at how MS-T compared to GVT for evaluating pronunciation samples. Overall, considering all phonological elements together, MS-T had significantly stronger correlations than GVT. However, the results were mitigated when we delved into each of the phonological criteria. The GVT and MS-T results showed varying degrees of correlation with human rater results across different aspects of pronunciation. Specifically, GVT had higher correlation for *comprehensibility*, *phonemes*, and *thought groups, sentence stress, and intonation*, while MS-T correlated more closely for *connected speech* and *word stress and rhythm*. This study examined whether a relationship existed between MS-T and human-rated scores with the goal of determining if this ASR platform could provide valid and reliable automated scoring of pronunciation that aligns with human judgements. The MS-T scores have been shown to be reliable and valid as the correlations between MS-T and human scores fare quite strong. As a result, this study demonstrates there is a possibility of MS-T implementation as an automated pronunciation scoring feature for proficiency tests, thereby improving the practicality of these tests, possibly leading to a reduction in costs. In addition, the results show that MS-T can potentially be deployed to automatically score pronunciation assessments, such as low-stakes language tests (proficiency placement tests). Further research is needed to confirm this perspective. In future research, it could be of interest to find out why these differences exist between the two ASR systems. Finally, the study shows that leveraging MS-T's ASR capabilities for L2 testing has the potential to provide a more efficient and cost-effective assessment tool compared to traditional language proficiency tests. It also suggests that MS-T can serve as a viable alternative to GVT for L2 pronunciation scoring.

Acknowledgements

We would like to express our gratitude to Carol Johnson for her guidance throughout this study. We also wish to thank Kathleen Brannen, Susan Jackson, Suzanne Springer, and Beau Zuercher, all from UQAM, for their assistance with the pronunciation samples; and Vicente Amati and Taras Kamtchatnikov, also from UQAM, for their technical support.

References

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, *27*(3), 355-377. https://doi.org/10.1177%2F0265532210364404

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, *73*(2), 113-123. https://doi.org/10.1093/elt/ccy055

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, *10*(3), 1-16. https://doi.org/10.1186/s40468-020-00101-6

Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 257-268). Springer.

Johnson, C., Cardoso, W., Zuercher, B., Brannen, K. et Springer, S. (2022). Using Google Voice Typing to automatically assess pronunciation. In *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (p. 203–207). Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2022.61.1459

Këpuska, V., & Bohouta, G. (2017). Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx). *Journal of Engineering Research and Applications*, *7*(03), 20-24.

McCrocklin, S., & Edalatishams, I. (2020). Revisiting popular speech recognition software for ESL speech. *TESOL Quarterly*, *54*(4), 1086-1097. https://doi.org/10.1002/tesq.3006

Mroz, A. (2020). Aiming for advanced intelligibility and proficiency using mobile ASR. *Journal of Second Language Pronunciation*, *6*(1), 12–38. https://doi.org/10.1075/jslp.18030.mro

Xerri, D., & Briffa, P. (Eds.). (2018). *Teacher involvement in high-stakes language testing*. Springer.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16972

A pilot study of data-driven learning approach in teaching Chinese vocabulary

Yu-Ting Tseng[a] and Li-Ping Chang[b]
[a]Graduate Program of Teaching Chinese as a Second Language, National Taiwan University, r07146012@ntu.edu.tw and [b]Graduate Program of Teaching Chinese as a Second Language, National Taiwan University, , lchang@ntu.edu.tw

How to cite: Tseng, Y.; Chang, L. (2023). A pilot study of data-driven learning approach in teaching Chinese vocabulary. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16972

Abstract

The Data-Driven Learning (DDL) approach advocates a shift from passive knowledge recipients to active researchers among learners. This is achieved by utilizing abundant and context-rich target language inputs in a bottom-up learning process (Johns, 1990). Despite the scarcity of empirical research on the implementation of DDL in Chinese as a Second Language (CSL) classrooms, this study conducted a teaching experiment focused on confusable words to explore the approach's effects and learners' attitudes. Five advanced-level CSL learners from diverse native language backgrounds participated in the study, being taught ten sets of confusable words over a five-week period. The first five lessons employed an indirect DDL method, while the latter five utilized a direct DDL method with Sketch Engine. To evaluate the effectiveness of the DDL approach, a pre-test was administered prior to the course, followed by a post-test, a questionnaire, and one-on-one interviews after the completion of the course. The results of the post-test revealed an average improvement of 24% compared to the pre-test with statistical significance. Additionally, learners exhibited a positive attitude towards the course, expressing a preference for learning vocabulary through collocations and showing a desire to observe pre-selected concordance lines under the guidance of the instructor.

Keywords: *data-driven learning, corpus, DDL instructional design, Chinese teaching, empirical study.*

1. Introduction

Johns (1990) proposed Data-Driven Learning (DDL), which advocates a bottom-up learning process driven by analyzing typical, large, and contextualized data of the target language. Numerous studies have validated the benefits of the DDL approach for students (Mizumoto & Chujo, 2015; Boulton & Cobb, 2017; Lee, Warschauer & Lee, 2019). However, its application in teaching Chinese as a Second/Foreign Language (CSL/CFL) has remained limited (Chang, 2022). Only Smith (2011) and Yeh & Zhang (2018) have implemented DDL in CFL classrooms, with the former incorporating five DDL tasks within textbook content, and the latter conducting an empirical study aimed at enhancing learners' usage of the discourse-linking connection *jiù* 'then' in oral storytelling. It is hoped that through this study, more CSL instructors will become aware of and be inclined to this approach in their teaching. Therefore, this study poses two primary research queries:

1. How effective is the application of DDL in enhancing learners' understanding and usage of confusable words in Chinese?

2. How do learners' preferences vary between direct and indirect DDL approaches, and what are their perspectives on instructional arrangements?

2. Method

This study conducted a five-week course consisting of ten sessions for two hours per week, while recruiting five learners with an advanced level of Chinese proficiency to participate. The first five sessions employed an indirect DDL (hands-off) approach, and the subsequent five sessions employed a direct DDL (hands-on) approach (Boulton, 2010).

To assess the effectiveness, learners were requested to undergo both a pre-test and a post-test, each comprising 40 multiple-choice vocabulary items. Upon completing all courses, they are required to fill out a questionnaire mainly adapted from Yoon and Hirvela (2004) and participate in one-on-one interviews.

3. DDL Instructional Design

3.1. Instructional Design for Indirect DDL Sessions: jiéguǒ 'result' and hòuguǒ 'consequence'

The key to the successful implementation of indirect DDL lies in the creation of paper-based materials for learners to observe. For illustration, we outline the steps involved in preparing materials focused on *jiéguǒ* 'result' and *hòuguǒ* 'consequence'. To commence, the instructor initiated an examination of the Chinese learner corpora, with the objective of scrutinizing and identifying instances of confusion prevalent among learners. The majority of these errors stemmed from disparities in the semantic prosody of the two words. 'Semantic prosody,' as elucidated by Louw (2000), refers to "a form of meaning which is established through the proximity of a consistent series of collocates." Specifically, the semantic prosody of *jiéguǒ* 'result' is neutral, while *hòuguǒ* 'consequence' always appears in a negative context. This finding was also indicated in the study of Xiao & McEnery (2006). Nevertheless, learners often employed *jiéguǒ* within a negative context, as exemplified in the following instance.

(1) *Rúguǒ wéile jiějué jī'è de wèntí, fàngqì lǜsèshípǐnde zhòngyàoxìng de huà, zhè zàochéng de ***jiéguǒ*** shì kěyǐshuō bùkānshèxiǎng. (If we abandon the importance of green food to solve the hunger problem, the consequences will be unthinkable.)

In addition, *jiéguǒ* has two usages, functioning as both a conjunction and a noun. Learners may mistakenly assume that *hòuguǒ* also has these two usages, leading to the production of error as shown in (2).

(2) *Lìngwài, wǒmen dìqiú de zīyuán yóuyú wǒmen méi kǎolǜ yīzhí shǐyòng, xiànzhì yě méiyǒu, ***hòuguǒ*** wǒmen miànlín quēfá zīyuán. (Moreover, we have not been mindful in utilizing our planet's resources, which are not unlimited. Consequently, we are faced with resource shortages.)

Six sheets for the instruction of *jiéguǒ* 'result' and *hòuguǒ* 'consequence' were prepared in the classroom. Each sheet contained ten sentences with a single focus. This aided in facilitating easy observation and analysis, ultimately enhancing learners' confidence. The selection of sentences in the corpus followed the criteria of typical usage contexts and functionalities. The order of the teaching materials began with the noun *hòuguǒ*, followed by the noun *jiéguǒ*, and then the conjunction *jiéguǒ* 'therefore'. It was presumed that learners learned the noun usage of *jiéguǒ* before while *hòuguǒ* is introduced as a new word. Hence, the presentation of *hòuguǒ* usage preceded the comparison with the usage of *jiéguǒ.* This sequence intended to first acquaint learners with the distinctions in collocating words when both *hòuguǒ* and *jiéguǒ* function as nouns. Subsequently, the focus shifted to the exclusive function of *jiéguǒ* as a conjunction, demonstrating a typical usage where it indicated an unexpected outcome. This was illustrated by employing the adverb *què* (however) to indicate a contrasting

function. The instructional materials were printed on a single side to facilitate easy comparison and reference for learners, enabling them to readily observe the key differences in usage between *hòuguǒ* and *jiéguǒ.*

In the classroom, the modified five-step instructional design was adopted from Smith (2011): Observation and Analysis, Hypothesis Formulation, Hypothesis Confirmation, Summarization, and Consolidation and Application. The aim was for students to observe, hypothesize, confirm, and revise their understanding of confusable words, followed by consolidation and application of their newfound knowledge. After several cycles of these steps, students were encouraged to summarize their findings, which were then reinforced through practical exercises to solidify and apply their newly acquired knowledge.

3.2. Activities Using Sketch Engine for direct DDL Sessions: zhíyè 'occupation' and hángyè 'industry'

The preparatory work for direct DDL is similar to the indirect method, the difference lies in the fact that instructors do not need to create paper-based materials; instead, they allow learners to directly operate corpora. In this study, the Sketch Engine was employed as the instructional tool (Kilgarriff et al., 2004). Before achieving the highest efficacy in the classroom, instructors must familiarize themselves with the usage of Sketch Engine and annotate each step to design classroom activities (Chang, 2022).

For illustration, we take *zhíyè* 'occupation' and *hángyè* 'industry' as examples. In the learner corpora, the authors observed that learners did not tend to misinterpret *zhíyè* as *hángyè*, but they did make the error of using *hángyè* in place of *zhíyè*, as shown in examples (3) and (4).

(3) *Wǒ cóng wǎnglùshàng kàndào nǐmen de zhāopìn, wǒ hěn yǒu xìngqù, tèbié duì dǎoyóu zhè yī ***hángyè***. (I saw your job advertisement on the internet and I am interested, particularly in the field of tour guiding.)

(4) *Yīncǐ, wǒ xīwàng néng yìngzhēng guǎnggào shèjìshī zhè yī fèn ***hángyè***, wéi guìgōngsī bànshì. (Hence, I aspire to secure the role of an advertisement designer in your esteemed company.)

In addition to this finding, the authors utilized the Corpus Of Contemporary Chinese (COCT) to further analyze the critical differences between the two words. Furthermore, the focus of instruction was placed on highlighting that *zhíyè* differed from *hángyè* in that, despite both being nouns, *zhíyè* possessed the additional function of being used as an attributive modifier meaning 'professional'.

In the classroom, the tasks assigned to students offered clear steps and objectives for their observations. Examples of tasks were as follows:

Task 1: Identify the three most frequent collocations for *zhíyè* and *hángyè*.

The activity involved guiding learners to use concordance functions to observe the usage of *zhíyè* and *hángyè* separately. The instructor employed questioning techniques to facilitate the learners' observations, such as "What are the three most common words that frequently appear together with *zhíyè*?" Similarly, learners were prompted to identify the three most frequently associated terms with *hángyè* and the typical words occurring on its left and right sides. The objective of the instruction was to enable learners to ultimately deduce that *zhíyè* can serve as an attributive modifier for nouns, such as 'athlete, actor, soldier, assassin,' etc., expressing the meaning of 'professional' or 'specialized'. On the other hand, *hángyè* did not possess such a usage.

Task 2: Use the Thesaurus Function to differentiate *zhíyè* and *hángyè.*

The second instructional activity involved guiding learners to access the Thesaurus Function and input *zhíyè* and *hángyè* separately. Through the utilization of visualizations and diagrams, learners were led to discover the semantic differences between the two terms. The innermost circle represented words closely related to key concepts, while the outer circles indicated more distant associations. As depicted in Figure 1, the meaning of *zhíyè* is closely related to terms such as skills, and profession. Figure 2 reveals that the terms closely associated with the semantic context of *hángyè* include sector, enterprise, field, and market, etc. By completing the tasks,

learners were guided towards concluding that *zhíyè* and *hángyè* referred to different meanings, and *zhíyè* can be an attributive often to modify the word on the right.

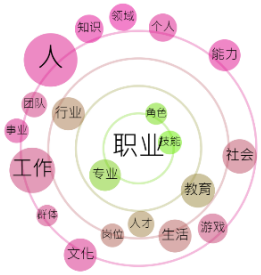


Figure 1. Thesaurus result of *zhíyè*

Figure 2. Thesaurus result of *hángyè*

4. Results

The average score for the pretest was 71%. The posttest had an overall average score of 95%, indicating a 24% improvement compared to the pretest. The Wilcoxon signed-rank test showed a statistically significant improvement in applying DDL to learn confusable words (p-value = 0.039).

The results of questionnaires and interviews were divided into three aspects. First, in the realm of learning preferences learners exhibited a pronounced inclination toward the concept of collocation, deeming it advantageous to their learning endeavors. Furthermore, they expressed a preference for the indirect approach due to its perceived impact on learning efficiency and its ability to bolster their confidence in the analysis of genuine linguistic data.

Second, in the evaluation of the course learners provided positive evaluations for it, and they perceived the guidance provided by the teacher in the classroom as highly significant. Learners placed significant emphasis on the pivotal roles of Hypothesis Formulation and Hypothesis Confirmation in both indirect and direct DDL methods. However, there existed varying opinions regarding the instructional arrangement of teaching a set of confusable words within a one-hour session, which may be attributed to the learners' individual learning strategies and personal learning styles.

Third, the experiences of learners using the tool and their interview responses can be summarized as follows. The majority of learners considered the corpus a valuable reference tool; however, learners did not achieve proficiency in utilizing the functions within Sketch Engine after five instructional sessions. It was suggested in the future to contemplate adopting an integrated approach to prevent learner fatigue and mitigate its impact on learning efficiency. For example, an integrated 50-minute class structure could encompass 30 minutes dedicated to indirect DDL, followed by 20 minutes allocated to direct DDL.

5. Conclusions

In light of learners' expressed preference for the indirect DDL, it is advisable for instructors aiming to implement DDL instruction to initiate their pedagogical endeavors with the indirect DDL approach. This initial step serves the purpose of acquainting learners with the five-step procedural framework, encompassing observation and analysis, hypothesis formulation, hypothesis revision, summarization, and consolidation and application. The arrangement can release learners' pressure to learn the corpus skills and data-driven steps simultaneously. Such an approach serves to enhance learners' proficiency and confidence in the analysis of linguistic concordance before progressing to the direct DDL method. Moreover, instructors should allocate ample time for constructive discussions and reflective activities to facilitate a comprehensive and enriched learning experience.

The study aimed to equip instructors with effective instructional design strategies for using DDL to teach Chinese confusable words, thereby enhancing the teaching and learning efficiency of these words. Despite the study's small sample size, the mitigation of this limitation was achieved through the application of a

comprehensive course design and rigorous research methodology. Based on the post-test results, it was evident that learners, following instruction on confusable word pairs, exhibited an average pass rate improvement of 24%. In light of these course outcomes and learners' feedback, this study recommends that educators explore the application of data-driven learning in various facets of language instruction, including vocabulary, grammar, and writing pedagogy.

Acknowledgements

We would like to thank the research support from the National Science and Technology Council (NSTC 112-2410-H-002-060), with assistance from Ms. Chun-ting Chou in conducting statistical significance tests.

References

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language learning*, *60*(3), 534-572.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393. https://doi.org/10.1111/lang.12224

Chang, Li-ping. (2022). The preliminary study of corpus literacy training for in-service Chinese language teachers. *Journal of Chinese Language Teaching*, *19*(4), 83-124.

Chujo, K., Anthony, L. and Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner- friendly, online parallel concordancer. In M. Mahlberg, V. González-Díaz, and C. Smith (Eds.), *Proceedings of 5th Corpus Linguistics Conference 2009*, University of Liverpool, UK. http://ucrel.lancs.ac.uk/publications/cl2009

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria Newsletter* (July 1990) (pp. 14-34).

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004, July 6-10). The Sketch Engine. Paper presented at *XI EURALEX International Congress*, Lorient, France.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721-753.

Louw, B. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer and H. Sauntson (Eds.), *Words in context: In honour of John Sinclair* (pp. 48-94). University of Birmingham, Birmingham.

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, *22*, 1-17.

Smith, S. (2011). Corpus-based tasks for learning Chinese: a data-driven approach. *The Asian Conference on Technology in the Classroom Official Conference Proceedings* 2011 (pp. 48-59).

Wang, P., Hsu, C., Long, S. & Liles, X. (2020). Designing Data-Driven Learning Activities for the Chinese as a Second Language Classroom. *Journal of Chinese Language Teaching*, *17*(3), 103-137.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103-129.

Yeh, M., & Zhang, X. (2018). Corpus-based instruction: Teaching discourse-linking jiu (就) in storytelling. *Chinese as a Second Language*, *53*(1), 1-23.

Yoon, H. & Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16865

Data-driven learning beyond English: Insights and implications from three monographs

Luciana Forti[a], Nina Vyatkina[b] and Eva Schaeffer-Lacroix[c]
[a]Department of Italian Language, Literature and Arts in the World, University for Foreigners of Perugia, , luciana.forti@unistrapg.it; [b]Department of Slavic, German, and Eurasian Studies, University of Kansas, , vyatkina@ku.edu and [c]Teacher Training Department, Sorbonne University, , eva.lacroix@sorbonne-universite.fr.

How to cite: Forti, L.; Vyatkina, N.; Schaeffer-Lacroix, E. (2023). Data-driven learning beyond English: Insights and implications from three books. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16865

Abstract

This paper stems from the 2023 CorpusCALL SIG Symposium on Data-Driven Learning (DDL) for Languages Other Than English (LOTE). Three monographs on DDL for LOTE were presented and are briefly illustrated in this paper (Forti, 2023; Vyatkina, 2024; Schaeffer-Lacroix, 2019). Despite the growth of DDL in second language (L2) research and education in many languages, its focus has largely been on English. As a result, the knowledge about the applicability of DDL for learning LOTE remains limited. This hinders the validity and generalizability of DDL as a whole, and conceals the important implications related to bridging the research vs. practice gap in such contexts where the focus is on LOTEs. The monographs introduced in this paper demonstrate the relevance of DDL for learning LOTEs by discussing corpus-based resources, pedagogical applications, and empirical research from three perspectives.

Keywords: *corpus linguistics, data-driven learning, German, Italian, languages other than English.*

1. Introduction

DDL is an increasingly popular field in language education and research. Corpus-informed materials for L2 English have grown significantly. However, LOTE resources remain scarce (Jablonkai et al., 2020). This disparity extends to research on the effectiveness of pedagogical corpus applications, where most studies focus on L2 English. Recent research synthesis on DDL reveals that 89% of empirical studies have been conducted in English teaching contexts (Boulton & Vyatkina, 2021). This limits the validity, generalizability, and practical applicability of research findings in multilingual educational settings, and may perpetuate an undesirable monolingual standard. To obtain a more complete understanding of the field and enhance its practical implications in multilingual pedagogical contexts, more corpus applications for LOTE are needed.

Here, the authors address this gap by discussing their monographs that represent first book-length publications entirely devoted to DDL for LOTE. By doing so, the authors report on multi-year projects with or learning scenarios for university students in three different linguistic and geographical settings. The authors advance the DDL field by discussing available DDL resources, pedagogical applications, and empirical research for Italian and German.

2. A review of three monographs on DDL for learning LOTE

2.1. Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning (Forti, 2023)

The use of corpora in Italian language pedagogy has attracted the interest of scholars for at least 30 years (Forti, 2023; Polezzi, 1993). However, only nine empirical studies over a total of 489 have focused on L2 Italian (Boulton & Vyatkina, 2021), and nearly all of them employed questionnaires to elicit learner attitudes and behaviors while engaging with DDL activities, with little insight into language gains. Furthermore, the corpus-based resources used were mostly unavailable to the public, as were the data elicitation tools, consequently hindering replicability of research.

The author's empirical study makes use of publicly available corpora and contains the data elicitation tools used. It was conducted in eight classes of (native language) L1 Chinese students, enrolled in a foundation language course at an Italian university. The eight classes were randomly assigned to either the experimental condition (DDL activities) or control condition (traditional activities). The experimental group consisted of 62 students (female = 47), with age range 18-27, while the control group consisted of 61 students (female = 38), with age range 18-26. A 1-hour lesson focused on verb-noun combinations was taught once a week in the eight classes for eight weeks. At the end of the lessons, a questionnaire, aimed at eliciting learner attitudes towards collocation learning and DDL, was administered. A test evaluating knowledge of the collocations was administered at four-week intervals. The last administration of the test was conducted four weeks after the end of the lessons to measure retention rates.

In terms of language gains, we found U-shaped learning curves in both groups, with no significant differences between them. This may by due to the fact that the length of the pedagogical intervention consisted of fewer than ten sessions: according to a meta-analysis of DDL studies (Lee et al., 2019), interventions with more than ten sessions tend to have a larger effect on language gains. We also identified a tendency towards higher retention rates in the DDL group. This indicates that the specific traits of concordance-based DDL activities (i.e. being able to sift through multiple contextualized examples of a word combination, and then detect regularities) are likely to produce longer-term learning. Semantic transparency, L1-L2 congruency, and dimensions of collocation knowledge were included in the analysis, providing a more nuanced view of the findings. In terms of learner attitudes, some initial difficulties were reported in relation to working with concordance lines. Nevertheless, they recognized the usefulness of the activities in enhancing their awareness about word combinations.

2.2. Corpus applications in language teaching and research: The case of data-driven learning of German (Vyatkina, 2024)

The systematic review part of the monograph provides an overview of the history of DDL for L2 German language and reveals a rich tradition dating back to the late 19th century, with frequency lists for German emerging in the 1920s. Various pedagogical applications for German, including the influential Routledge Frequency Dictionary series, have since evolved. L2 German DDL resources encompass different corpus interaction methods (direct/indirect, hands-on/hands-off), publication formats (reference books, textbooks, tests, monographs, articles), and linguistic areas (vocabulary, grammar, lexico-grammar, pragmatics). While L2 German DDL research is smaller in scale compared to L2 English, it stands as a prominent LOTE target. Numerous L2 German studies (Vyatkina, 2024, Chapter 3) offer detailed insights into DDL implementations, showcasing their effectiveness when learners receive ample support. However, the field calls for increased methodological diversity, a broader range of targeted language skills, consistent reporting, and long-term studies.

The focus in the empirical study part of the monograph was on efficacy of teaching L2 German collocations to U.S. university students by combining Instructed Second Language Acquisition (ISLA) and DDL approaches. The data were collected from one intact group of high-intermediate proficiency learners in an L2 German course who were exposed to both a paper-based and a computer-based DDL treatment and consented to participate in the study. The study compared pretest-posttest gains in lexical, morphological, and collocational knowledge. The results confirmed the effectiveness of explicit interventions in developing productive L2 collocation knowledge, aligning with the usage-based theory of language acquisition (e.g. O'Keeffe, 2021). The study also validated DDL as an

effective method for teaching collocations, with both DDL methods bringing significant knowledge gains with an advantage of the computer-based method for morphological knowledge and because of its efficiency. This research extended the scope of DDL beyond English to inflectional languages, demonstrating its applicability to both lexical and grammatical collocations. It emphasized the integration of best practices from ISLA and DDL research and encouraged cross-disciplinary collaboration (O'Keeffe, 2021).

The pedagogical applications part of the monograph addresses how open-access corpora can serve as Open Educational Resources (OERs). 'Incorporating Corpora' (Vyatkina, 2020) is presented as one such OER, which is tailored for English-speaking L2 German learners and teachers. It utilizes DWDS (Digitales Wörterbuch der Deutschen Sprache), an open access German corpus and tool suite, offering interactive online exercises linked to DWDS. This OER includes user-friendly tutorials and instructions for both educators and learners, addressing key DDL issues while adhering to web accessibility guidelines (Meunier, 2022). Successful pilot testing with intermediate-level students that affirmed its instructional value is reported.

2.3. Encounters with German language in use promoted by pedagogical corpora (Schaeffer-Lacroix, 2019)

As mentioned in section 2.1, German is the best represented foreign language in DDL research after English. According to Boulton and Vyatkina's (2021) meta-study on DDL articles written in English, eight out of sixteen articles concern L2 German and are designed for learners whose L1 is English; only two papers represent German L2 learners whose L1 is French. If one expands the list of DDL research on German L2 for learners whose L1 is French to studies written in French, more results can be identified: two reports on the creation of corpora for learning L2 German in France (Trouvain et al., 2013; Wigham & Poudat, 2020), ten out of fourteen DDL studies conducted by Schaeffer-Lacroix [1], and her research presented in this paper.

Here, the author's monograph presents a brief overview of five corpus-based scenarios designed for teaching German in France in different learning settings. Their scientific background is inspired by researchers such as Bachelard, Bruner, Chanquoy, Sweller, Tricot, Pekarek Doehler, Piaget, and Vygotsky. Pedagogical concepts like discovery learning, intertextual text production, interaction between expert and novice, and language awareness structure its five scenarios, which cover the whole range between low and high instruction levels. Learners' perceptions of the effectiveness of DDL were identified with the help of filmed interviews led by an external researcher right after the classroom experiments set up for three out of the five scenarios (for details, see Schaeffer-Lacroix 2016, 2018b, 2022 in the list provided in footnote 1). These perceptions were compared to the learners' activities tracked through filmed computer screens and audio recordings of pair discussions. The author defends the general idea that small, specialized corpora like those created for her five learning scenarios can support DDL research and practice in an effective way, even if, for statistical reasons, Dodd (1997) considers those datasets which contain fewer than one million tokens do not merit the label 'corpus'. However, in the eyes of foreign language learners, corpora are big enough if they help them find answers to their questions, and students can even be discouraged by a huge amount of data (Schaeffer-Lacroix, 2009, p. 194 & 200). Crosthwaite and Baisa (2023) warn the DDL community not to ignore the growing importance of artificial intelligence applications, and they recommend turning them into partners instead of considering them as a threat for language education. This is one more reason to use small corpora whose content and quality can be (semi-)manually checked; this makes them relevant for a given task, be it for particular language learning needs and contexts or as training material for machine learning.

3. Discussion and conclusions

Overall, these three monographs offer new insights and resources for researchers, language teaching practitioners, and students interested in corpus-based learning and teaching methods (Table 1). It appears that opening the floor for LOTE, not only with respect to the target language but also with respect to the language of publication, offers the opportunity to share the knowledge and the methods stemming from different cultures, to renew the DDL field

[1] DDL research on German as a foreign language

and to reaffirm its importance at a time when its existence its is challenged by artificial intelligence applications.

Table 1. Summary of the monographs.

	Forti (2023)	Vyatkina (2024)	Schaeffer-Lacroix (2019)
Target language	Italian	German	German
Teaching context	Italian L2 language courses for prospective university students in Italy.	German as a foreign language courses in the USA.	German as a foreign language courses in France.
Main objective	To illustrate the main methodological challenges in researching DDL effects and to demonstrate, by means of an empirical study, how such challenges may be addressed.	To address three existing divides in the DDL field between: 1) English and other languages; 2) DDL and ISLA research; and 3) research and pedagogical practice.	To inform on how to integrate DDL activities in language learning scenarios.
Systematic review	DDL teaching materials and research studies.	DDL teaching materials and research studies.	Comparison of DDL activities to other CALL activities. Available corpora and corpus tools.
Empirical study	Comparing the effect of DDL vs. non-DDL activities aimed at developing phraseological competence.	Comparing the effectiveness of hands-on and hands-off DDL for teaching verb-noun collocations.	The effect of DDL on language awareness (prepositions, verb particles, commas) and on learning to write according to the constraints of the given text genre.
Pedagogical applications	Principles and resources to develop DDL activites.	A suite of open access, interactive DDL activities.	Five teaching scenarios.

Publishing research in LOTE supports multilingualism and allows DDL researchers to stick closely to their audience. Using another language than English may moreover help with designing activities inspired by a non-English cultures and avoid biases introduced by the English-language perspective. The authors support the strengthening of the theoretical grounding of DDL research, the integration of pedagogical DDL applications with more learner-friendly user interfaces, and the enhancement of rigor in study design and reporting.

References

Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds), *Teaching and learning by doing corpus analysis* (pp. 9–26). Rodopi B.V.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology*, *25*(3), 66–89. https://www.lltjournal.org/item/10125-73450

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 100066. https://doi.org/10.1016/j.acorp.2023.100066

Dodd, B. (1997). Exploiting a corpus of written German for advanced language learning. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds), *Teaching and language corpora* (pp. 131–145). Longman.

Forti, L. (2023). *Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning*. Routledge.

Jablonkai, R.R., Forti, L., Castelló, M. A., Iguenane, I. S., Schaeffer-Lacroix, E., Vyatkina, N. (2020). Data-driven learning for languages other than English: the cases of French, German, Italian, and Spanish. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thouësny (Eds), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 132-137). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.48.1177

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Meunier, F. (2022). Revamping DDL: Affordances of digital technology. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 344–360). Routledge.

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching*, *54*, 259–272. https://doi.org/10.1017/S0261444820000245

Polezzi, L. (1993). Concordancing and the teaching of ab initio Italian language for specific purposes. *ReCALL*, *5*(09), 14–18. https://doi.org/10.1017/S0958344000004067

Schaeffer-Lacroix, E. (2009). *Corpus numériques et production écrite en langue étrangère. Une recherche avec des apprenants d'allemand* [Electronic corpora and learning to write in a foreign language] [PhD thesis, Sorbonne nouvelle]. HAL. https://theses.hal.science/tel-00439095

Schaeffer-Lacroix, E. (2019). *Corpus et didactique de l'allemand – La langue à bras-le-corps* [Encounters with German language in use promoted by pedagogical corpora]. Lambert-Lucas.

Trouvain, J., Laprie, Y., Möbius, B., Andreeva, B., Colotte, V., Fauth, C., Fohr, D., Mella, O., Jügler, J., & Zimmerer, F. (2013). Designing a bilingual speech corpus for French and German language learners. *Proceedings of Corpora and Tools in Linguistics, Languages, and Speech*, 32–34. https://www.coli.uni-saarland.de/~trouvain/Trouvain_et_al_2013.pdf

Vyatkina, N. (Ed.). (2020). *Incorporating corpora: Using corpora to teach German to English-speaking learners* [Online instructional materials]. University of Kansas, Open Language Resource Center. https://corpora.ku.edu

Vyatkina, N. (2024). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge.

Wigham, C. R., & Poudat, C. (2020). Corpus complexes et standards : Un retour sur le projet CoMeRe [Complex corpora and standards: a review of the CoMeRe project]. *Corpus*, *20*. https://doi.org/10.4000/corpus.4736.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16989

Japanese readability assessment using machine learning

Tyler Ivie[a] and Robert Reynolds[b]
[a]Department of Linguistics, Brigham Young University, tyler.j.ivie@gmail.com and [b]Office of Digital Humanities, Brigham Young University, robert_reynolds@byu.edu

How to cite: Ivie, T.; Reynolds, R. (2023). Japanese readability assessment using machine learning. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16989

Abstract

We present a new corpus of Japanese texts, labeled according to six second-language readability levels. We also show the results of experiments training machine-learning classifiers to automatically label new texts according to reading level. The resulting models can be used in language-learning websites and applications to enhance Japanese language learning. The best-performing model, Random Forest, achieved an F1 score of 0.86, with an adjacent accuracy of 0.97. Of the 114 features used, we identify a small subset of five features that are sufficient to achieve an F1 score of 0.74. The corpus, code, and resulting models are free and open-source.[1]

Keywords: *readability, machine learning, Japanese.*

1. Introduction

The goal of assigning reading levels to texts has been approached through a variety of methods. Readability metrics and subsequent classifications have been produced using simple formulas, such as syllable to word ratio as seen with the Flesch Reading Ease scale (Flesch, 1948), statistical analysis (e.g. Lee & Hasebe, 2016), and machine-learning (e.g. Hancke et al., 2012).

All methods have their unique merit, and the results of each are distinct. Formulaic approaches are simple enough to be computed by hand, providing a helpful insight and often being more language-agnostic (Bendová & Cinková, 2021). The Flesch scale, for example, was used for predicting the readability of Czech. Statistical approaches allow far more factors to be considered without being too cumbersome and can produce very useful results, albeit statistical models require more expertise to execute than a formula, but they can be automated and even made available online (Hasebe & Lee, 2015).

Machine-learning approaches are much like statistical models but can be fed additional features to consider, giving them the potential to be even more versatile, whereas formulaic approaches like the Flesch scale, considering the least amount of features, are most easily tricked. Many formulas have been created for the English readability assessment over the years to overcome genre-specific and demographic-specific classification challenges (Klare, 1974), where machine-learning models, considering more features, can perform well in more varied contexts. Traditional machine learning models cannot easily be adapted between languages, however, and are time-consuming to create.

[1]https://github.com/reynoldsnlp/japanese_readability_corpus

In our search we were not able to find examples where machine-learning was implemented to classify the readability of Japanese documents for L2 readers. Most current literature regarding the automated reading level assessment of Japanese leverages statistical models involving features carefully selected from large corpora (Lee & Hasebe, 2016), often using features like the proportion of different Japanese scripts and length of sentences. There is also record of non-automated, social approaches to document classification, where different versions of the same document are given to test-groups of native speakers to collect human-reported difficulty metrics (Sakai, 2011).

It is our aim to help expand resources available to Japanese language learners and professionals with machine-learning readability classification using open-source software and freely available corpora. The code used to obtain and analyze our corpus will be included in links to our Github repository along with a simple script to predict the grade level of texts from outside of the training dataset. Most of the corpus used in this paper is in the creative commons and is provided in our codebase, but for one to replicate the results of the study, part of the corpus will need to be downloaded from the publisher.

2. Method

The labeled corpus was created by collecting all of the freely available graded readers from two websites: tadoku.org and jgrpg-sakura.com. Each document was given a level 0-6, 0 being the simplest and 6 being the most difficult. Levels 1-5 correspond to JLPT levels 5-1, whereas level 0 in our corpus represents a below JLPT N5 level reader. A sampling of academic articles and news stories were also added manually. Image count was also manually collected but was not considered in the final machine-learning model.

Document text from tadoku.org was stripped from graded-reader PDFs. Furigana, pronunciation guide characters, were removed systematically by comparing font size ratios. Text was manually cleaned and audited to ensure at least 98% accuracy. Image-only PDFs were digitized using Tesseract optical character recognition. In a few cases purely manual work was necessary because of poor implementation of vertical text in some file formats.

Documents from jgrpg-sakura.com were downloaded in .html format and parsed, removing <ruby></ruby> tags to ensure no furigana were swaying the data. These texts were also manually reviewed but required less rigorous auditing due to the less extensive and non-manual edits made to the documents. Because jgrpg-sakura.com uses a 9-level system as opposed to the 6-level system used by jreadability.net and tadoku.org, individual documents from jgrpg-sakura.com were assigned a 6-level equivalent beforehand, following a brief manual review.

Academic articles and news were assumed to be native-level and were assigned level 5. They received the least amount of processing, containing no furigana. Characters that triggered unresolved escaped in, however, namely backslashes and ascii control characters were removed. The final document count in this small labeled corpus is 167.

For machine learning, we extracted 114 linguistic features from each document. Stanza (Qi et al., 2020), a python library for natural-language-processing was used for part-of-speech and basic dependency tagging. The features extracted consist of basic orthographic features (e.g. ratio of hiragana, katakana, and kanji): vocabulary and frequency features (e.g. proportion of words at a certain level on the Japanese Language Proficiency Test (JLPT)); and grammatical features (e.g. particle count, certain conjugations, and a sampling of specific morpho-syntactic patterns).

Token frequency lists were taken from http://corpus.leeds.ac.uk/frqc/internet-jp.num and kanji frequency lists were sourced from https://scriptin.github.io/kanji-frequency/wikipedia/. Since official JLPT vocabulary lists

have not been posted since the Japanese government added a fifth level, JLPT N3 in 2010,[2] vocabulary level lists were sourced from https://tangorin.com/vocabulary/, one of many informal but researched resources for assigning JLPT levels to Japanese vocabulary.

After features were extracted, we trained models using 5-fold cross-validation, including the following algorithms from Scikit Learn (Pedregosa et al, 2011) and xgboost (Chen & Guestrin, 2016): XGBoost (XGB), XGBoost Random Forest (XGBRF), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Classification And Regression Trees (CART), Naïve Bayes (NB), and Support Vector Machine (SVM).

Later, importances were generated and used in tandem with recursive feature elimination in another Random Forest model to track the performance of the corpus RF model with and without different features in order to determine the features that were most important.

3. Results

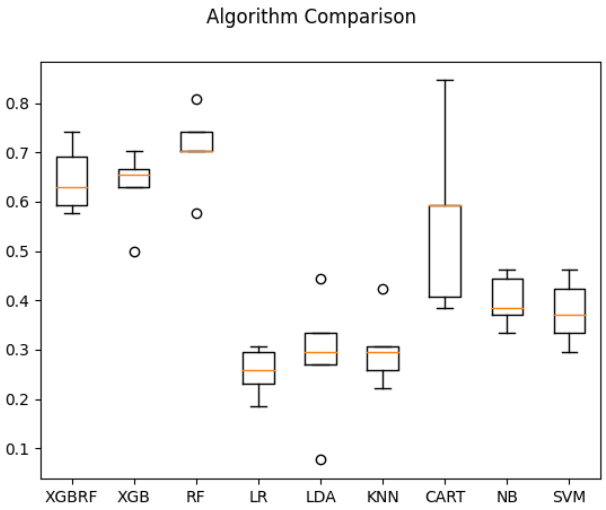


Figure 1. F1 scores on 5-fold cross-validation.

As seen in Figure 1, the Random Forests classifier was found to consistently outperform other models, so further evaluation focuses on this model. A full classification report for the Random Forest model is given in Table 1. Precision measures the accuracy of positive predictions, while recall measures the completeness of positive predictions. The F1-score is the harmonic mean of precision and recall. Table 1 shows that the average F1 score for all levels is 0.86. The lowest performance was for levels 1 and 2.

[2] https://www.jlpt.jp/e/topics/list2010.html

	Precision	Recall	F1-score	Support
Level 0	0.88	0.88	0.88	8
Level 1	0.57	0.80	0.67	5
Level 2	1.00	0.60	0.75	5
Level 3	0.86	0.86	0.86	7
Level 4	1.00	1.00	1.00	4
Level 5	1.00	1.00	1.00	5
Accuracy	---	---	0.85	34
Macro Avg	0.88	0.86	0.86	34
Weighted Avg	0.88	0.85	0.86	34

Table 1. Classification report for Random Forest model on 5-fold cross-validation

The confusion matrix in Table 2 shows where individual predictions of the model relate to their actual readability levels. Gray cells along the diagonal are correct predictions. This shows that almost all of the Random Forest model's mistakes are only off by one level, yielding an adjacent accuracy score of 0.97. This indicates that the model is not only making good predictions in general, but that even its mistakes are reasonably close to the actual readability level.

		Predicted Level					
		Lv 0	**Lv 1**	**Lv 2**	**Lv 3**	**Lv 4**	**Lv 5**
Actual Level	**Lv 0**	7	1	-	-	-	-
	Lv 1	1	4	-	-	-	-
	Lv 2	-	1	3	1	-	-
	Lv 3	-	1	-	6	-	-
	Lv 4	-	-	-	-	4	-
	Lv 5	-	-	-	-	-	5

Table 2. Confusion matrix for Random Forest model on 5-fold cross-validation

To evaluate which features are most important to the model, we used Recursive Feature Elimination (RFE) to rank features. Then, to determine the minimum number of features needed for a viable model, we iteratively trained models on more and more features, using the order from RFE.

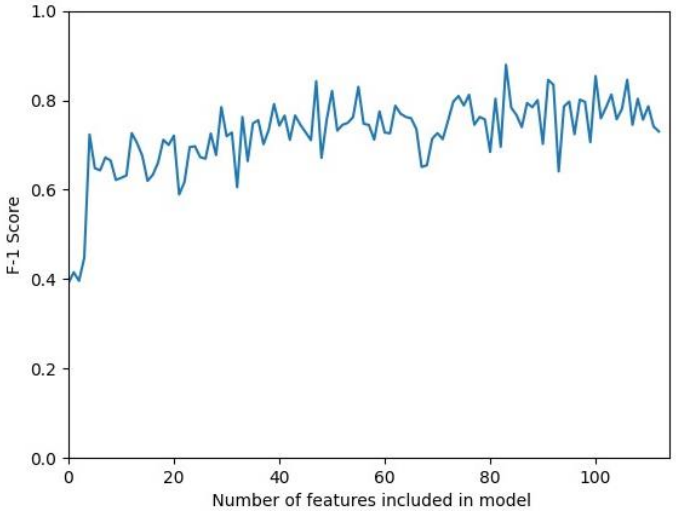


Figure 2. F1 score of models trained with top-N number of features, ranked by importance

The results in Figure 2 show that reasonable performance can be achieved with only five features: type-token ratio, type-lemma ratio, kanji type-token ratio, unique verbs to token ratio, and mean sentence length. These five features represent three different categories of features: lexical variation, orthographic variation, and syntactic complexity.

4. Conclusions

We have collected a small second-language readability corpus of Japanese with 34 documents labeled for six readability levels. We also trained a Random Forest classifier that achieves an average F1 score of 0.86 on cross-validation. This classifier can be implemented in websites and applications to support Japanese language learning.

This study is limited by a small corpus size, but the consistent performance of the Random Forest model, especially its near-perfect adjacent accuracy, suggests that the textual features included in the model are valid identifiers of readability at these six reading levels. In particular, the five features with the highest importance are crucial to identifying the readability of Japanese texts: type-token ratio, type-lemma ratio, kanji type-token ratio, unique verbs to token ratio, and mean sentence length. Especially noteworthy is the importance of orthographic variation, which is almost completely absent from readability research with other languages.

Future work is needed to build a larger corpus to increase confidence in the findings reported here. Although many commercial resources exist, we intend to focus on adding texts with licenses that allow publishing the corpus with an open license.

To our knowledge, this is the first published research using a machine-learning approach to automate Japanese readability classification. Despite limited resources, the results are quite promising, and future work in this domain is likely to see significant gains in automated readability classification of Japanese texts.

Acknowledgements

We are grateful to Brigham Young University's College of Humanities for funding this research.

References

Bendová, K., & Cinková, S. (2021). Adaptation of classic readability metrics to czech. Paper presented at the *International Conference on Text, Speech, and Dialogue*, 159-171.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Paper presented at the *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785-794.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233. https://doi.org/10.1037/h0057532

Hancke, J., Vajjala, S., & Meurers, D. (2012). Readability classification for german using lexical, syntactic, and morphological features. Paper presented at the *Proceedings of COLING 2012*, 1063-1080.

Hasebe, Y., & Lee, J. (2015). Introducing a readability evaluation system for japanese language education. Paper presented at the *Proceedings of the 6th International Conference on Computer Assisted Systems for Teaching & Learning Japanese*, 19-22.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, , 62-102.

Lee, J., & Hasebe, Y. (2016). Readability measurement of japanese texts based on levelled corpora. *The Japanese Language from an empiricaL perspecTive*, 143.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020, July). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101-108).

Reynolds, R. (2016). Insights from russian second language readability classification: Complexity-dependent training requirements, and feature evaluation of multiple categories. Paper presented at the *Proceedings of the 11th Workshop on Innovative use of NLP for Building Educational Applications*, 289-300.

Sakai, Y. (2011). Improvement and evaluation of readability of Japanese health information texts: an experiment on the ease of reading and understanding written texts on disease. *Library and information science*, (65), 1-35.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16959

Learner corpora and the design of data-driven learning activities

Luciana Forti[a]
[a]Department of Italian Language, Literature and Arts in the World, University for Foreigners of Perugia, luciana.forti@unistrapg.it

How to cite: Forti, L. (2023). Learner corpora and the design of data-driven learning activities. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16959

Abstract

This paper seeks to continue the conversation on an underused resource in Data-Driven Learning (DDL), namely learner corpora. It first explores the potential of learner corpora in data-driven learning activity design, focusing on the advantages claimed by a number of scholars throughout the years and mainly associated with the field of Learner Corpus Research (LCR). It then illustrates the status that learner corpus use has in DDL activity design, on the basis of data drawn from the most recent and comprehensive review of DDL research, covering a timespan of 30 years. After describing the main qualitative and quantitative design features of a learner corpus of Italian (i.e. the CELI corpus*), the paper shows how learner corpora containing texts produced at different proficiency levels can be used as graded corpora for both target-oriented and error-oriented activities. The sample illustrated activities can thus cater to learners at different proficiency levels, including lower-intermediate and intermediate levels, which are still under-represented in DDL research. Some of the main pedagogical and scientific advantages of using level-specific DDL materials in a paper-based format are also presented and briefly discussed.*

Keywords: *Learner corpora, data-driven learning, Common European Framework of Reference for Languages, Italian.*

1. Introduction

Learner corpora have seen major developments in the past few decades. These developments have concerned not only the techniques used to build them, their accessibility, and the learner languages they represent, but also the methodological sophistication with which the linguistic data they contain is analysed. We, thus, have increasingly greater insight into interlanguage features and dynamics, and this can invaluably enhance our understanding of second language acquisition theories. Learner corpus use remains, however, quite limited in pedagogical settings, such as those involving Data-Driven Learning (DDL).

In this paper, we review the potential and status of learner corpus use in DDL, showing how there is a gap between calls for increased learner corpus use in DDL and its actual use in DDL activities. We then introduce and describe the CELI corpus, a learner corpus of Italian, containing written texts produced in the context of language certification exams across four balanced proficiency levels (from B1 to C2). Finally, we present some ways in which a learner corpus such as the CELI can be used in designing DDL activities, and then briefly discuss some of the related advantages.

2. Potential and status of learner corpora in DDL

In DDL, "[…] the task of the language teacher is to provide a context in which the learner can develop strategies for discovery - strategies through which he or she can 'learn how to learn'" (Johns, 1991, p. 1). While originally DDL developed mainly with reference to the use of L1 corpora, the adjacent and partially overlapping field of Learner Corpus Research (LCR) has argued that even "[…] learner corpora can be extremely useful for form-focused instruction, because they present students with typical interlanguage features" (Gilquin & Granger, 2022, p. 433). In particular, learner corpora can provide data on typical errors, as well as examples representing good language use in cases where no errors are present. In these latter cases, learner corpora with good examples of the target language can be particularly useful when there is a need to match the level of language difficulty to a certain proficiency level.

When used in conjunction with a reference L1 corpus (e.g. Ackerley, 2017), the learner corpus plays an important role in the observation of over- and under-use phenomena, the detection of false friends, the detection of non-idiomatic expressions and the identification of errors. By means of systematic and guided comparison between the two corpora, learners are placed within a context in which they have the tools to discover differences between learner-produced and L1-speaker-produced language. This comparison can be even more effective when the learner corpus is a *local* learner corpus (Seidlhofer, 2002). Local learner corpora contain the texts produced by the same learners who will then analyse them in the context of DDL activities. The advantage of this kind of corpus exploration is that learners can analyse features of their own interlanguage through the various data types offered by a corpus (concordances, frequency lists, dispersion graphs, etc.), and this can provide a considerable boost for motivation, since the learners are actively involved in creating a corpus from their own pieces of writing. The many potential and advantageous uses of learner corpora in DDL are very much present in the literature and in the conference presentations, especially in contributions from recent years (Gilquin, 2023; Götz, 2022).

But to what extent are learner corpora actually used in DDL? To answer the question, we can filter the data contained in Boulton & Vyatkina (2021)'s 30-year review on DDL research according to the type of corpus used. The findings show that only 21 of the total of 489 papers collected by the authors involve the use of a learner corpus, which corresponds to a mere 4.29%. Furthermore, in most cases, the learner corpus seems to be a local learner corpus and its use is implemented in conjunction with a reference corpus.

We will now describe the features of an off-the-shelf learner corpus of Italian, and its consequential potential uses in DDL, in the hope of further shedding light on the usefulness of learner corpora in DDL activity design, despite their limited presence in DDL research and pedagogy so far.

3. The CELI corpus

The CELI corpus (Spina et al., 2022, 2023)[1] is a learner corpus of Italian. It consists of written texts collected from the Italian language certification exams known as CELI, *Certificati di Lingua Italiana*[2]. The texts were collected from passed language certifications exams at levels B1, B2, C1, and C2. Texts from lower levels of proficiency were not collected as they were very few, very brief, and the product of highly guided tasks, which would have made it difficult to isolate the actual independent contribution of the learner. Each text, originally written on paper, was manually transcribed on a digital platform. Formal aspects which could have potentially hindered the quality of the post-tagging were normalised. The total number of texts is 3,041, which amount to about 600,000tokens. Each proficiency level contains a comparable quantity of tokens. The nationality backgrounds of the test subjects are varied, with Greek, Spanish, and Romanian being the three most represented ones. The search interface allows selection of not only the proficiency level and the nationality background, but also the age group (from 10-14 to

[1] The corpus can be accessed at the following two webpages: https://lt.eurac.edu/cqpweb/, https://apps.unistrapg.it/cqpweb/. (last accessed: 31/07/2023).

[2] The CELI exams are developed and administered all around the world by the CVCL – Centro per la Valutazione e le Certificazioni LInguistiche, based at the University for Foreigners of Perugia. More information about the CELI language certification system may be found at the following webpage: https://www.unistrapg.it/node/457 (last accessed: 31/07/2023).

75-79), the exam centre location (Abroad/Italy), the task number, the sex, the text genre, and the text type. The CELI is one of the very few learner corpora balanced according to proficiency level, and in which text attribution to a certain level is based on a sound framework, which in this case is provided by the *Common European Framework of References for Languages* (CEFR) (Forti, 2023). Table 1 summarises the main qualitative design features of the CELI, while Table 2 summarises its main quantitative features.

Table 1. Qualitative design features of the CELI corpus (based on classification by Tono, 2003).

Language-related features	Task-related features	Learner-related features
Mode: written.	Data collection: pseudo-longitudinal.	Internal-cognitive: age. Internal-affective (motivation/attitude): n/a.
Genre: mixed (article, blog, e-mail, essay, letter, report, story).	Elicitation: passed language certification exams.	L1 background: by approximation (i.e. nationality; n. 104, mostly Greek, Spanish, Romanian)
Style: mixed (argumentative, descriptive, narrative and mixed).	Use of references: no.	L2 environment: L2/FL.
Topic mixed (leisure, current affaires, etc.)	Time limitation: fixed.	L2 proficiency: yes, based on certification exam.

Table 2. Quantitative features of the CELI corpus

CEFR level	Number of texts/learners	Number of tokens
B1	1,212	156,612
B2	840	152,251
C1	585	149,859
C2	404	149,892
TOTAL	3,041	608,614

4. Using the CELI corpus for graded DDL activities

This section contains some examples of how a balanced learner corpus, based on the CEFR, can provide useful data for both target-oriented and error-oriented DDL activities. The former may be designed as guided-discovery activities, aimed at guiding the learners toward the observation of form-meaning patterns in the concordance lines. The latter may be designed to aim for the learner to detect the error within a set of concordance lines. In both cases, the examples within the concordance lines which learners will engage with will be suitable for the particular

proficiency level at which they are.

One area where learners of L2 Italian struggle with is *andare* ('to go') + PREPOSITION | ARTICLE + NOUN costruction. Most errors are found in the choice of the preposition and/or article that needs to be used between the verb and the noun. By extracting the data related to this construction from the CELI corpus, we are able to obtain good examples of error-free sentences containing this construction, as well as errors concerning the construction. Appendix 1 provides a table containing a line of concordance lines extracted from the CELI corpus. The table is divided into four sections, one for each proficiency level. Within each proficiency level, we see a section with good examples and one with an erroneous example. With the former, we may ask the learners to try and detect any regularities that are typical when the target construction is used. Each learner will be able to engage with concordance lines that are suitable for their level. The example containing the error, on the other hand, may be used in activities where the learners are presented with a set of concordance lines, within which one example contains the error. The task for the learners will be to identify which concordance line contains the error. In trying to reach this goal, the learners will apply the meta-cognitive strategies that are typical of DDL, such as scanning, comparing, and making hypotheses.

Other kinds of activities can be developed, such as matching multiple split sentences or filling the gap in a provided set of concordance lines. In each of these cases, the activities may be uploaded online by using one of the many applications that are available and suitable for this purpose. However, these activities can also be paper-based (Boulton, 2010). This choice carries with it at least four major advantages: 1) once the activities have been created, they can be re-used indefinitely, so the initial time spent developing them will be time saved in the future; 2) though relying on data extracted from corpora, these activities will not require the use of computers on the learners' part, and this can be particularly useful with large groups or in contexts where access to computers may still be challenging; 3) the fact that concordance lines are pre-selected by a teacher means that even corpora that do not have a learner-friendly interface can be used for DDL activities; and 4) re-usable DDL activities facilitate replication in empirical investigations aimed at evaluating the effects of DDL in terms of language gains, learner and teacher attitudes, and learning processes that they activate.

5. Final remarks and conclusions

In this short paper, we aimed at continuing the conversation on the potential of learner corpora in DDL activities. After illustrating the potential and status of learner corpus use in DDL research, we described the main design features of the CELI corpus, a learner corpus of Italian. We then briefly presented some ideas on how a corpus such as the CELI corpus may be used in DDL.

The potential of a learner corpus in DDL activity development is inextricably linked to its design. The variety of designs on which learner corpora are based determines not only the potential but also the limitations that a learner corpus will have with regard to its applicability to DDL activity development. A corpus containing balanced subcorpora of different proficiency levels has the specific advantage of potentially catering for learners at those different levels, thus offering the opportunity to have graded input both in relation to good examples of language, as well as in relation to the typical errors that may be found in the language production of learners at different proficiency levels.

We look forward to further discussion on the topic of learner corpus use in DDL activity design and its many connections to other relevant topics in the field, such as the need for grading DDL activities, the pedagogical and scientific advantages of paper-based DDL materials, and the 'learner-friendliness' of corpus-based resources.

References

Ackerley, K. (2017). Effects of Corpus-Based Instruction on Phraseology in Learner English. *Language Learning & Technology*, *21*(3), 195–216.

Boulton, A. (2010). Data-Driven Learning: Taking the Computer Out of the Equation: Data-Driven Learning. *Language Learning*, *60*(3), 534–572.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning and Technology*, *25*(3), 66–89.

Forti, L. (2023). *Exploring the affordances of CEFR-based learner corpora in Data-driven learning* [Plenary presentation], Japan Association for English Corpus Studies (JAECS) Spring Forum 2023, JAECS SIG on DDL, online, May 13, 2023.

Gilquin, G. (2023). Written learner corpora to inform teaching. In R. R. Jablonkai & E. Csomay (Eds), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (pp. 281–295).

Gilquin, G., & Granger, S. (2022). Using data-driven learning in language teaching. In A. O'Keeffe & M. J. McCarthy, *The Routledge Handbook of Corpus Linguistics* (2nd ed., pp. 430–442). Routledge. https://doi.org/10.4324/9780367076399-30

Götz, S. (2022). *Learner corpora and DDL: A Promising Synergy?* [Paper presentation in Symposium], EUROCALL CorpusCALL SIG symposium, Online.

Johns, T. (1991). Should you be persuaded – two examples of data-driven learning materials. *Classroom Concordancing*, *English Language Research Journal 4*, 1–16.

Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (Vol. 6, pp. 213–234). John Benjamins. https://doi.org/10.1075/lllt.6.14sei

Spina, S., Fioravanti, I., Forti, L., Santucci, V., Scerra, A., & Zanda, F. (2022). Il Corpus CELI: Una nuova risorsa per studiare l'acquisizione dell'italiano L2. *Italiano LinguaDue*, *14*(1), 116–138. https://doi.org/10.54103/2037-3597/18161

Spina, S., Fioravanti, I., Forti, L., & Zanda, F. (2023, in print). The CELI Corpus: Design and linguistic annotation of a new online learner corpus. *Second Language Research*. https://doi.org/10.1177/02676583231176370

Tono, Y. (2003). Learner corpora: Design, development and applications. *Paper Presented at the Corpus Linguistics 2003 Conference (CL 2003) Lancaster*, 800–809.

Appendix 1. CELI corpus data for cline of DDL activities focused on the *andare* ('to go') + PREPOSITION | ARTICLE + NOUN costruction.

B1		
Target-oriented examples	Prima abbiamo partecipato alla cerimonia e dopo **siamo andati al ristorante**. **Sono andata nel parco** vicino alla mia università. Adesso posso **andare al lavoro** in bicicletta e non devo andare a piedi. Nel tempo libero, mi piace **andare nel centro storico**. Penso **di andare alla fiera** anche il prossimo anno.	Main features: short sentences; literal meanings.
Error-oriented example	Sono **andata sulla cerimonia** di premiazione.	Error type: wrong selection of preposition.
B2		
Target-oriented examples	Ho deciso di **andare all'estero** per studiare. Mi è venuta una grande depressione, e insomma, sono dovuta **andare dallo psichiatra e dalla psicologa** per aiuto. Il giorno dopo la separazione **sono andata al lavoro**, ma c'era solo il mio corpo. Tutti i lunedì **andavamo al mercato** insieme. Avevo deciso di sposarla, in futuro. Ho visitato la città di Bangkok e **sono andato sull'isola** Kho Tao.	Main features: literal meanings.
Error-oriented example	Ho provato a continuare le cose che amavo fare, **come andare nelle montagne**, incontrare gli amici.	Error type: wrong selection of preposition.
C1		
Target-oriented examples	Una foto qui un'altra là e **via vanno la privacy e il contenuto** delle storie dietro una foto. Abbiamo preso il tram **e siamo andate ai magazzini**. Abbiamo deciso con mio marito di **andare alla laguna** dei sette colori che si chiama Bacalar in Messico. La mia lettera successiva **andrà al Presidente**. Per commentare la gente doveva scrivere su un foglio, **andare nell'ufficio postale** e finalmente inviare la lettera.	Main features: some non-literal meanings.
Error-oriented example	Puoi **andare sui passi** di Dracula ed incontrarlo magari nei tuoi sogni.	Error type: non-idiomatic expression.
C2		
Target-oriented examples	Si rese conto che doveva spiegare alla sua miglior amica **come erano andate le cose** l' altra sera . Non appiattarsi **e lasciarsi andare all'ozio** ma smuovere le risorse del cervello. **Siamo andati dal Giudice**, io e l' altro autista. Avevo deciso di **andarci col treno** nonostante i numerosi cambiamenti tra Genova e la Sicilia. Con tutte le tasse che stiamo pagando e una percentuale dovrebbe **andare alle casalinghe**.	Main features: literal and non-literal meanings; idiomatic expressions.
Error-oriented example	Questa nuova abitudine di fidarti di un algoritmo automatizzato piuttosto che **andare l'avventura** - sia amicale o romantica - nella vita reale.	Error type: error in idiomatic expression.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16952

Making ICALL's intelligence accessible and understandable: evaluation, validation, and future directions of ICALL ecosystems

Jasper Degraeuwe[a] and Patrick Goethals[b]
[a]Department of Translation, Communication and Interpretation, Ghent University, , Jasper.Degraeuwe@UGent.be and
[b]Department of Translation, Communication and Interpretation, Ghent University, , Patrick.Goethals@UGent.be

How to cite: Degraeuwe, J.; Goethals, P. (2023). ICALL ecosystems: evaluation, validation and future directions. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16952

Abstract

In this paper, we will reflect upon the revised and expanded version of the Intelligent Computer-Assisted Language Learning (ICALL) ecosystem developed in Degraeuwe and Goethals (2022a). We portray the upgraded version of the ecosystem and perform a reflective analysis on it following the same methodology as presented in Degraeuwe and Goethals (2022a). The analysis is carried out by means of a case study with 22 participants who study Spanish as a foreign language and who work with the ecosystem in a blended writing course focused on vocabulary learning and lexical ambiguity. Students' attitudes towards engaging in the ICALL ecosystem are gauged through a questionnaire, which revealed a slight but non-significant positive change in attitude (compared to a statistically significant positive change in Degraeuwe and Goethals 2022a). However, it should be noted that the initial attitudes before engaging with the ecosystem were already very positive (5.32 on an 8-point Likert scale), which might have made it more challenging for the ecosystem to have a considerable impact. Additionally, an analysis at question level showed that sparking curiosity about language technology and providing a good user experience remain challenging aspects.

Keywords: *Intelligent Computer-Assisted Language Learning, Natural Language Processing, reflective analysis, user attitudes.*

1. Introduction

With applications such as example sentence selection systems (Pilán et al., 2016), difficulty classifiers (Tack, 2021) and exercise generation tools (Bodnar, 2022; Heck & Meurers, 2022), the integration of Intelligent Computer-Assisted Language Learning (ICALL) in language learning courses can be a valuable addition to the arsenal of teaching methods, for example as a complement to on-campus vocabulary learning activities (Ruiz et al., 2021). Nevertheless, using ICALL also comes with its limitations. Recognising lexically ambiguous items and distinguishing between their senses is one of those pending issues in ICALL research, especially for languages other than English (Degraeuwe & Goethals, 2022b). In fact, the Natural Language Processing (NLP) driven technique of Word Sense Disambiguation (WSD) is rarely integrated in corpus query tools or in the development of computer readable resources for didactic purposes, such as graded word lists (Tack, 2021). Additionally, if end users (which may refer to students, teachers or even course book designers) want to fully exploit the potential of ICALL environments, they should possess certain technological metaskills that allow them to decide which queries are relevant and feasible to perform and which are not (Schweinberger, 2021).

In Degraeuwe and Goethals (2022a), we analysed how both abovementioned aspects could be addressed by designing an "ICALL ecosystem" (Section 2.1). Working in such ecosystem led to a statistically significant positive change in attitude towards ICALL, but this did not automatically mean that the users also enjoyed working with the computer more, or that the ecosystem sparked their interest in learning more about language technology. With the present study, we aim to corroborate the findings of Degraeuwe and Goethals (2022a) by performing a similar case study based on a revised and expanded version of the ecosystem (see Section 2.3). Furthermore, this follow-up study will also briefly touch upon the use of generative Artificial Intelligence (AI), as the adequate and efficient use of AI driven applications is bound to become an essential skill in education.

2. Method

2.1. Ecosystem design

As described in Degraeuwe and Goethals (2022a), the ecosystem enables users to generate and use customised learning materials **(Aspect_1)** and helps them gain technological metaskills by stimulating their curiosity and promoting their autonomy **(Aspect_2)**. In the meantime, all user activities are saved in a structured database, which can then be used for improving the NLP driven methods integrated into the environment **(Aspect_3)**.

2.2. Case study design

The study follows the same design as Degraeuwe and Goethals (2022a). The 22 participants (Dutch-speaking students of Spanish as a foreign language enrolled in a 3rd bachelor B2+ level writing course at university) work with the online learning environment of the Spanish Corpus Annotation Project [1] (SCAP) (Goethals, 2018), which includes a corpus consultation component, exercise generator and collaborative section for research purposes.

Part of the course consists of completing a blended vocabulary learning module, which encompasses five hours of on campus classes and an online module on lexical ambiguity. During the classes, the students learn to use the corpus consultation and vocabulary learning functionalities of the SCAP platform **(Aspect_1)**. For the online module, they consider lexical ambiguity from the perspective of the computer by watching knowledge clips [2] and develop their own WSD models by making interactive exercises on lexically ambiguous vocabulary items in the collaborative section of the platform **(Aspect_2)**. The responses to those exercises are collected in a database and used to develop the actual WSD method integrated into the environment **(Aspect_3)**.

2.3. Modified aspects of the ecosystem

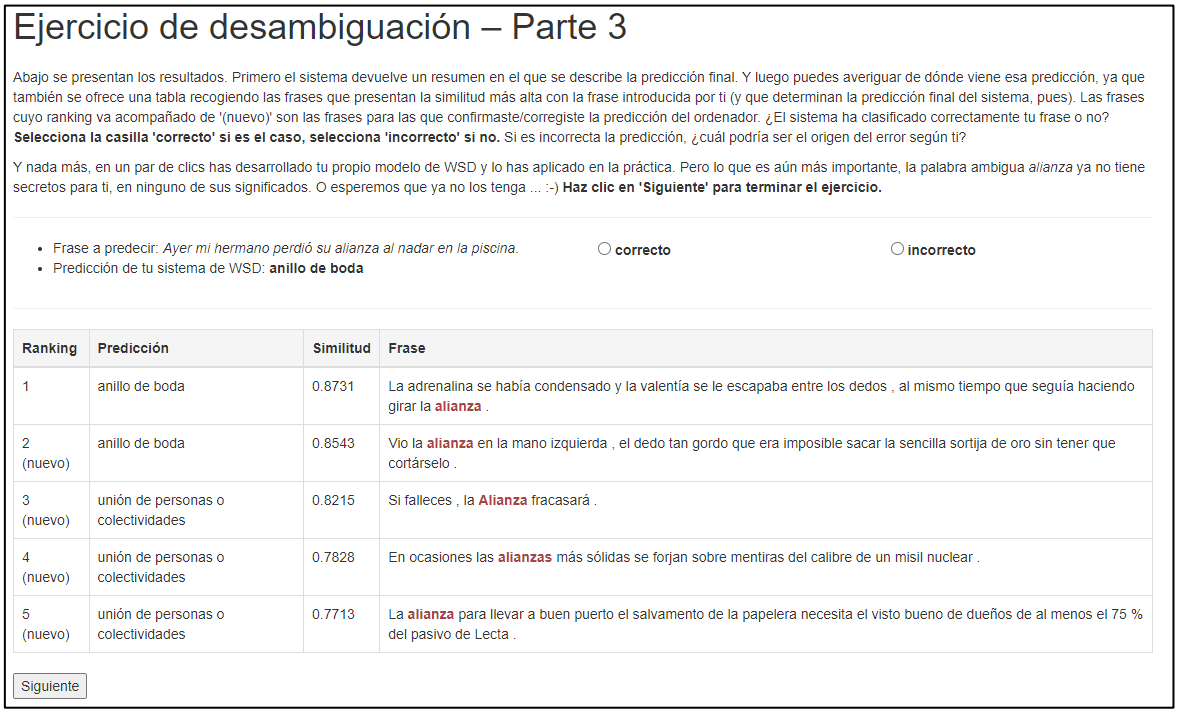
From Degraeuwe and Goethals (2022a) it could be concluded that, although the ecosystem led to enhanced insights into NLP and increased confidence in the computer as a learning assistant, the students did not necessarily enjoy working with the computer more or wanted to learn more about language technology. To address these shortcomings, we made the following adjustments to the ecosystem:

- Integration of short quizzes into the knowledge clips
- Addition of a free text component in the interactive exercises on lexical ambiguity (see Figure 1)
- Addition of "odd one(s) out" as a new type of exercise in the collaborative section of the platform: in this exercise, students are presented with six sentences for one particular meaning of a lexically ambiguous word after which they have to identify the sentence(s) which do not belong to that particular meaning

[1] Publicly available version of the platform accessible through scap.ugent.be. Demo video of the in-house version available at https://www.youtube.com/watch?v=RFaIWEEZcVM.

[2] Complete video available at https://www.youtube.com/watch?v=-ev56uEpIkA.

- Possibility to download customised learning materials created with NLP models (e.g. the WSD models) trained on the exercise response data collected in the ecosystem



Ejercicio de desambiguación – Parte 3

Abajo se presentan los resultados. Primero el sistema devuelve un resumen en el que se describe la predicción final. Y luego puedes averiguar de dónde viene esa predicción, ya que también se ofrece una tabla recogiendo las frases que presentan la similitud más alta con la frase introducida por ti (y que determinan la predicción final del sistema, pues). Las frases cuyo ranking va acompañado de '(nuevo)' son las frases para las que confirmaste/corregiste la predicción del ordenador. ¿El sistema ha clasificado correctamente tu frase o no? **Selecciona la casilla 'correcto' si es el caso, selecciona 'incorrecto' si no.** Si es incorrecta la predicción, ¿cuál podría ser el origen del error según ti?

Y nada más, en un par de clics has desarrollado tu propio modelo de WSD y lo has aplicado en la práctica. Pero lo que es aún más importante, la palabra ambigua *alianza* ya no tiene secretos para ti, en ninguno de sus significados. O esperemos que ya no los tenga ... :-) **Haz clic en 'Siguiente' para terminar el ejercicio.**

- Frase a predecir: *Ayer mi hermano perdió su alianza al nadar en la piscina.* ○ **correcto** ○ **incorrecto**
- Predicción de tu sistema de WSD: **anillo de boda**

Ranking	Predicción	Similitud	Frase
1	anillo de boda	0.8731	La adrenalina se había condensado y la valentía se le escapaba entre los dedos , al mismo tiempo que seguía haciendo girar la **alianza** .
2 (nuevo)	anillo de boda	0.8543	Vio la **alianza** en la mano izquierda , el dedo tan gordo que era imposible sacar la sencilla sortija de oro sin tener que cortárselo .
3 (nuevo)	unión de personas o colectividades	0.8215	Si falleces , la **Alianza** fracasará .
4 (nuevo)	unión de personas o colectividades	0.7828	En ocasiones las **alianzas** más sólidas se forjan sobre mentiras del calibre de un misil nuclear .
5 (nuevo)	unión de personas o colectividades	0.7713	La **alianza** para llevar a buen puerto el salvamento de la papelera necesita el visto bueno de dueños de al menos el 75 % del pasivo de Lecta .

Siguiente

Figure 1. Illustration of free text component added to exercise in which users can develop, use, and evaluate their own WSD models (for a translation, see Appendix 1). The table at the bottom presents the similarity values based on which the model obtained its prediction ("Predicción de tu sistema de WSD") for the sentence introduced by the user ("Frase a predecir")

2.4. Questionnaire

To gain insights into the potential of the upgraded ecosystem, the students are administered the same questionnaire as in Degraeuwe and Goethals (2022a), namely an adapted version of the Attitude towards CALL questionnaire (A-CALL; Vandewaetere & Desmet, 2009). The questionnaire contains 15 eight-point Likert scale questions (each of them representing a specific attitude towards ICALL; see Table 1) and is filled out by the students before and after completing the vocabulary learning module. Additionally, we include an extra multiple-choice question to analyse which technology assisted tools the participants use in their language learning process (see Figure 2 at the end of Section 3 for the multiple choice options).

3. Results

3.1. Attitudes towards ICALL

Table 1 reports the mean scores and Standard Deviation (SD) values of the 15 questions, with the scores from Degraeuwe and Goethals (2022a), hereafter referred to as the 'previous case study', being added between parentheses. The overall pre- and post-scores show a small positive change in attitude of 0.08, which did not appear to be statistically significant (paired samples t-test, p=0.45). However, it should be highlighted that the pre-scores were already quite elevated, which might have made it more challenging for the ecosystem to significantly affect students' attitudes. In fact, the overall pre-score of 5.32 more or less equals the overall post-score of 5.36 obtained in the previous case study. In summary, the overall results indicate that the ecosystem and its upgrades are not able to further improve attitudes which are already fairly to very positive before the interaction with the ecosystem takes place.

When breaking down the results at question level, the same tendency as in the previous case study can be observed: working in the ecosystem enhances students' insights into NLP (questions 1, 5 and 9) and increases their confidence in the computer as a vocabulary learning assistant (4 and 6), but this does not necessarily go hand in hand with an increased curiosity (2) and a better user experience (11). This tension also appears from the statistical significance at question level (paired samples t-test, $p<0.05$): the increases for questions 1 and 6 are significant, the decreases for questions 2 and 11 as well.

Table 1. Questionnaire results, with questions for which the scores were reversed being marked with $^{(rev)}$ and post-means with a statistically significant difference being marked with *.

Nr	Question	Pre (n=22)		Post (n=22)	
		Mean	SD	Mean	SD
1	The computer is able to analyse the grammatical characteristics of words and link words to their corresponding part of speech (noun, verb, adjective, etc.).	5.95 (5.12)	1.21 (1.62)	6.64* (7.18)	1 (0.86)
2	I am interested in knowing more about the technology which enables computers to automatically create vocabulary exercises and resources.	5.32 (4.16)	1.52 (2.2)	4.73* (3.89)	2.07 (1.87)
3	The computer only sees sequences of letters which are combined into words, it is not able to see meanings and concepts behind these sequences of letters.$^{(rev)}$	4.91 (4.91)	1.69 (1.51)	4.68 (5.96)	1.81 (1.48)
4	I have confidence in computer created vocabulary exercises and tests.	5.27 (4.69)	1.28 (1.31)	5.33 (5.86)	1.53 (1.3)
5	If I introduce a large collections of texts on a certain domain into a specific application, I think that this application will be able to return a keyword list with the most typical words for the domain.	6.41 (5.56)	1.3 (1.37)	6.68 (6.71)	0.95 (1.05)
6	The computer is able to generate vocabulary exercises and resources tailored to my proficiency level.	5.5 (5.34)	1.5 (1.21)	6.55* (6.68)	1.06 (1.22)
7	The teacher's attitude and enthusiasm towards and knowledge of computer-assisted vocabulary learning determine to a large extent my attitude towards using computers for vocabulary learning purposes.$^{(rev)}$	3.09 (3.22)	1.41 (1.77)	3 (3.54)	1.69 (2.12)
8	Computer-assisted vocabulary learning offers more flexibility to learning vocabulary in Spanish.	5.68 (5.33)	0.84 (1.5)	6.05 (5.61)	1.17 (1.89)
9	The computer is able to analyse the syntactic structure of sentences, and assign the correct syntactic function (subject, direct object, etc.) to words.	5.32 (4.53)	1.32 (1.27)	5.64 (5.61)	1.33 (1.47)
10	Computer assisted vocabulary learning is as valuable as traditional methods for vocabulary learning in Spanish.	5.77 (4.28)	1.48 (1.49)	5.95 (4.82)	1.43 (1.49)

11	I (would) like to learn Spanish vocabulary with the help of the computer.	6.77 (5.28)	1.07 (2.1)	5.95* (4.89)	1.21 (1.91)
12	I find it easier to accept an error committed by a language teacher than an error committed by the computer.(rev)	3.64 (4.06)	1.92 (1.78)	4 (4.07)	1.83 (2.02)
13	People who learn Spanish vocabulary through computer assisted learning methods are less proficient in Spanish than people who learn Spanish vocabulary through traditional paper and pencil methods.(rev)	6.5 (6)	1.19 (1.93)	6.32 (6.29)	1.52 (1.41)
14	Computer assisted vocabulary learning is a valuable extension of traditional learning methods for vocabulary learning in Spanish.	6.73 (6.09)	1.08 (1.47)	6.59 (6.07)	1.1 (1.74)
15	Vocabulary exercises and resources created automatically by an application cannot contain errors.(rev)	2.91 (3.34)	1.57 (2.06)	2.82 (3.18)	1.59 (1.93)
		5.32 (4.81)	**1.36 (1.64)**	**5.4 (5.36)**	**1.42 (1.58)**

3.2. Use of technology-assisted tools

Finally, we briefly discuss the results of the additional multiple-choice question on the use of technology-assisted tools by the students (Figure 2). The bar chart shows that language learning applications and machine translation systems are widely used by students, and CALL environments to a lesser extent (only 45% of the participants had experience with them before the start of the course). As for the use of generative AI, the results clearly show that AI-driven tools will inevitably become a part of language learning: before taking the course, none of the students had used tools like ChatGPT (the course started 3 months after the release of GPT-3.5), but by the end of the course this number had already risen to 8 of the 22 participants. In other words, this finding suggests that the integration of generative AI will become an important aspect of future ICALL research.

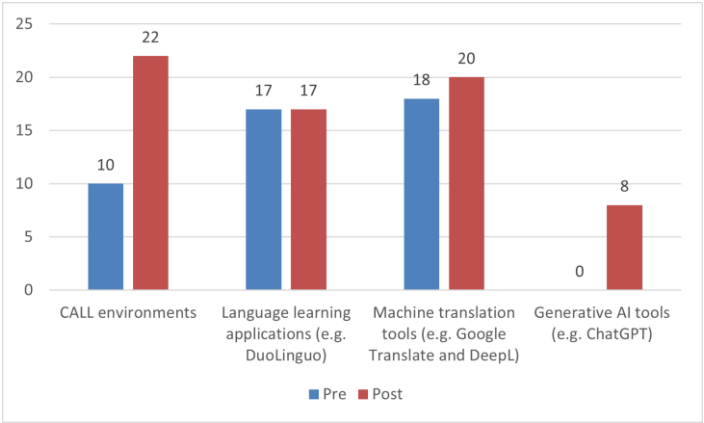


Figure 2. Bar chart presenting which technology-assisted tools are used by the 22 students

4. Discussion and conclusions

In this paper, we presented and reflected upon the upgraded version of the "ICALL ecosystem" introduced in Degraeuwe and Goethals (2022a). By using the same case study design we aimed to corroborate the findings of Degraeuwe and Goethals (2022a) which reported a significantly positive change in students' attitudes towards ICALL after having engaged with the ecosystem. However, this year's questionnaire results revealed only a small

(and statistically non-significant) positive change in attitude. It does have to be highlighted, though, that the initial attitudes of this year's participants were already very positive (average score of 5.32 on an 8-point scale), which might have made it more challenging for the ecosystem to achieve the same impact (note that last year's post-score amounted to 5.36, coming from a pre-score of 4.81). Additionally, despite the upgrades integrated into the ecosystem the findings again underlined the area of tension between what students consider to be the value, quality, and/or potential of learning methods and the user experience these methods provide/the interest they spark. Finally, the additional question on the use of technology assisted tools showed that these kinds of applications are widely used by students, with generative AI as an emerging new source.

To conclude, finding alternative ways to improve the user experience (e.g. the integration of generative AI) will require further research, although the lack of corroboration for the ecosystem's positive impact on user attitudes might also be considered an incentive to first redesign some of its core aspects. Furthermore, we also plan to perform experiments with teachers as a new type of target audience, for example by asking them to prepare a vocabulary learning class using the ecosystem. Finally, grouping the questionnaire items into different variables (e.g. curiosity and quality of learning assistance) and performing a correlation analysis on them could be another interesting avenue for future research.

Acknowledgements

This research has been carried out as part of a PhD fellowship on the IVESS project (file number 11D3921N), funded by the Research Foundation – Flanders (FWO). Additionally, we want to express our gratitude to the reviewers for their valuable feedback, and to Dr. Koen Plevoets for providing support on the statistical analysis.

References

Bodnar, S. (2022). The instructional effectiveness of automatically generated exercises for learning French grammatical gender: Preliminary results. In D. Alfter, E. Volodina, T. François, P. Desmet, F. Cornillie, A. Jönsson, & E. Rennes (Eds.), *Proceedings of the 11th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL 2022)* (Vol. 190, pp. 10–22). https://doi.org/10.3384/ecp190002

Degraeuwe, J., & Goethals, P. (2022a). ICALL ecosystems: making ICALL's intelligence both accessible and understandable. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds.), *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022* (pp. 89–94). https://doi.org/10.14705/rpnet.2022.61.1440

Degraeuwe, J., & Goethals, P. (2022b). Interactive word sense disambiguation in foreign language learning. In D. Alfter, E. Volodina, T. François, P. Desmet, F. Cornillie, A. Jönsson, & E. Rennes (Eds.), *Proceedings of the 11th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL 2022)* (Vol. 190, pp. 46–54). https://doi.org/10.3384/ecp190005

Goethals, P. (2018). Customizing vocabulary learning for advanced learners of Spanish. In T. Read, B. Sedano Cuevas, & S. Montaner-Villalba (Eds.), *Technological innovation for specialized linguistic domains: Languages for digital lives and cultures, proceedings of TISLID'18* (pp. 229–240). Éditions Universitaires Européennes.

Heck, T., & Meurers, D. (2022). Generating and authoring high-variability exercises from authentic texts. In D. Alfter, E. Volodina, T. François, P. Desmet, F. Cornillie, A. Jönsson, & E. Rennes (Eds.), *Proceedings of the 11th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL 2022)* (Vol. 190, pp. 61–71). https://doi.org/10.3384/ecp190007

Pilán, I., Volodina, E., & Borin, L. (2016). Candidate sentence selection for language learning exercises: From a comprehensive framework to an empirical evaluation. *Revue Traitement Automatique Des Langues*, *57*(3), 67–91.

Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, *25*(4), 510–539. https://doi.org/10.1177/1362168819872859

Schweinberger, M. (2021). *An evaluation of computational, corpus-based approaches to language learning* [Webinar]. International Perspectives on Corpus Technology for Language Learning - Seminar Series, University of Queensland, Australia. https://languages-cultures.uq.edu.au/event/session/7415

Tack, A. 2021. *Mark My Words! On the Automated Prediction of Lexical Difficulty for Foreign Language Readers*. Louvain-la-Neuve, Belgium: UCLouvain & KU Leuven PhD thesis

Vandewaetere, M., & Desmet, P. (2009). Introducing psychometrical validation of questionnaires in CALL research: The case of measuring attitude towards CALL. *Computer Assisted Language Learning*, *22*(4), 349–380. https://doi.org/10.1080/09588220903186547

Appendix 1. Translation of Figure 1

Below the results are presented. First, the system gives back an overview in which the final prediction is described. And then you can check where this prediction comes from because a table is added containing the sentences that show the highest similarity with the sentence introduced by you (and which thus determine the final prediction of the system). The sentences whose ranking is accompanied by '(new)' are the sentences for which you confirmed/corrected the prediction of the computer. Did the system classify your sentence correctly or not? **Tick the box 'correcto' if it's the case, tick 'incorrecto' if not.** If the prediction is incorrect, what could have caused the error according to you?

And that's it, in just a few clicks you've developed your own WSD model, and put it into practice. But what's even more important, the ambiguous item *alianza* doesn't hold any further secrets from you, in any of its meanings. Or let's hope that it doesn't … :-) **Click on 'Siguiente' to finish the exercise.**

- Sentence to predict: *Yesterday my brother lost his wedding ring while swimming in the pool*
- Prediction of your WSD system: **wedding ring**

Ranking	Prediction	Similarity	Sentence
1	wedding ring	0.8731	The adrenaline had been condensed and courage slipped through his/her fingers as he/she kept spinning the wedding ring.
2 (new)	wedding ring	0.8543	He/she saw the wedding ring on the left hand, the finger being so fat it was impossible to take off the simple gold ring without having to cut it off.
3 (new)	union of people or groups	0.8215	If you die, the Alliance will fail.
4 (new)	union of people or groups	0.7828	Sometimes the most solid alliances are forged based on lies with the calibre of a nuclear missile.
5 (new)	union of people or groups	0.7713	The alliance for the rescue of the paper business needs the approval of owners of at least 75% of Lecta's liabilities.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16966

Developing LEMI: A new corpus based literacy support tool for schoolchildren

Roxana Rogobete[a], Alexandru Oravițan[b], Mădălina Chitez[c] and Karla Csürös[d]

[a]Department of Romanian Studies, West University of Timisoara, roxana.rogobete@e-uvt.ro; [b]Department of Modern Languages and Literatures, West University of Timisoara, alexandru.oravitan@e-uvt.ro; [c]Department of Modern Languages and Literatures, West University of Timisoara, madalina.chitez@e-uvt.ro and [d]Department of Modern Languages and Literatures, West University of Timisoara, karla.csuros@e-uvt.ro

How to cite: Rogobete, R.; Oravițan, A; Chitez, M.; Csürös, K. (2023). Developing LEMI: A new corpus based literacy support tool for schoolchildren. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16966

Abstract

This study centres on the developing and testing stages of a literacy support tool dedicated to young schoolchildren. The LEMI tool is currently in development (since January 2023) at the CODHUS research centre (Centre for Corpus Related Digital Approaches to Humanities) from the West University of Timişoara, Romania. LEMI aims to stimulate interest in reading during the first individual and collective reading activities (ages 7-11). This aim will be achieved by creating a digital reading repository with a user-friendly interface that verifies reading text complexity and delivers automatic reading-level reports to users. We use corpus linguistics methods to create a text complexity formula adapted to the Romanian language, which can be integrated into the automated complexity evaluation interface in LEMI. The necessity of such an instrument is motivated by the fact that, in Romania, there are increased rates of functional illiteracy and school dropout. We hypothesise that texts must be level-adapted (according to grade or readability) for schoolchildren to relate positively to reading activities. In the Romanian context, LEMI is the first digital tool wholly tailored to children's literature, which complements national curricula and didactic materials provided to young children. Distinctively, LEMI responds to the need for easy-to-use tools to adapt reading individually, according to the reader's profile. LEMI is a unique tool, not only for Romanian but also for children's literature in other languages. The functionalities of the LEMI pilot version will be tested with the partners involved in the project (three schools from Timiș county and an educational NGO).

Keywords: *LEMI, literacy support tool, corpus based literacy tool, children's literature repository, text complexity automatic assessment, text complexity in Romania, readability for Romanian language.*

1. Introduction

Understanding the process of reading comprehension and developing reading proficiency is essential from a variety of perspectives, from the cognitive to the pedagogical and cultural approach. According to the OECD (2009, 2), reading literacy is defined as "understanding, using, reflecting on, and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society." Students' engagement is a key aspect of lifelong learning (Ho & Lau 2018), yet motivating children to pursue reading has proven to be a challenge in recent decades, as the digital transformation has become a commonplace.

The last PISA reports (n.d.) focused on 15-year-old students, and the assessment of their key knowledge and skills show an almost negligible decrease in the students' reading skills from 2015 to 2018 (all countries taken into consideration, see Figure 1).

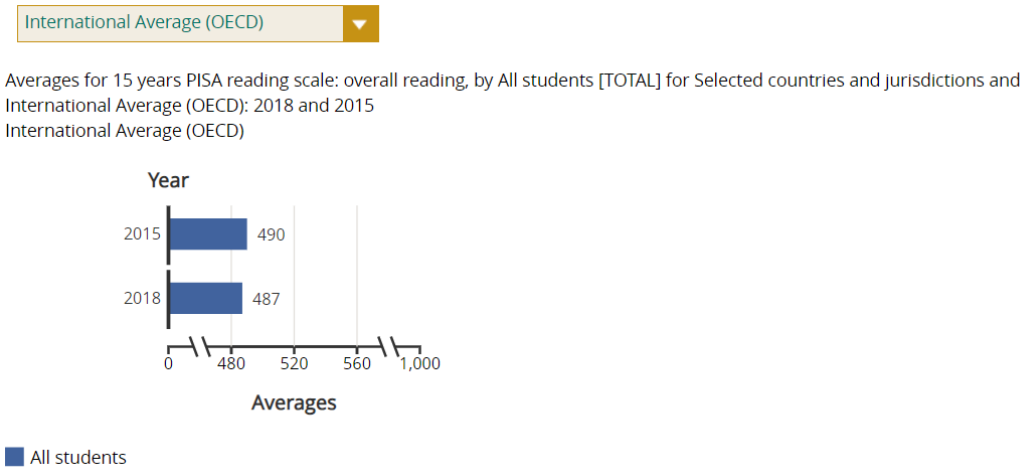


Figure 1. PISA reading scale: overall reading (All students, all countries).
https://pisadataexplorer.oecd.org/ide/idepisa/report.aspx

However, when the same criteria are selected for a country such as Romania alone, the scores are lower in all the subjects discussed (reading, mathematics, science, see OECD 2019, 1). Not only are the numbers lower in the case of Romanian students (with approximately 11-12%), but the decline is also faster (Figure 2): the reading attainment scores are overall lower for Romania and the difference between 2018 and 2015 is greater than the average for all students (as seen in Figure 1).

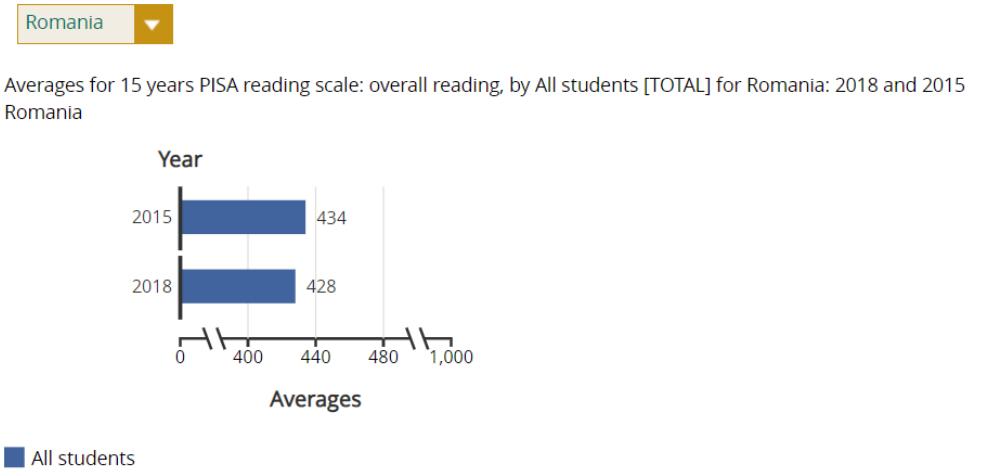


Figure 2. PISA reading scale: overall reading (All students, Romania).
https://pisadataexplorer.oecd.org/ide/idepisa/report.aspx

Over the last decade, Romania has seen an increase in functional illiteracy and school drop-out rates; the country has the highest rate of early school leavers in Europe (15%, compared to less than 10% at the European level, see Eurostat 2022, 1; and European Commission, 2022, 3), despite the fact that numerous initiatives have been taken to improve reading behaviour in both the 'traditional' and digital environments. Thus, the school curriculum assumes a direct link between reading activities and increasing the motivation to read. However, the lack of a rigorous way of structuring information based on the criterion of cognitive progression and systematicity

produces the opposite effect: decreased motivation to read. Despite the wide range of school textbooks approved by the National Center for Policies and Evaluation in Education (NCPEE), there are no unified school reading recommendations based on linguistic research, which have measured the complexity of the recommended texts (which in many cases have remained unchanged for decades) by reference to reading level (readability, in the technical literature) specific to each age category. Readability is defined as the degree to which a given group of people find a certain piece of reading comprehensible (McLaughlin, 1969) or as the ease with which a text can be understood due to its style of writing (Klare, 1963). In this context, there is a need to develop digital resources and tools tailored to the specific demands of distinct group ages or school levels. Although educational curricula and selected texts present in textbooks have to be related with research regarding measurement of text complexity, few children's reading apps have been developed, and the majority is limited to English language texts (see Chitez et al., 2023). In Romania, there are no school reading series whose content has been assessed at the level of text complexity, but there is one reading application that contains digital versions of literary texts. However, they are selected using human recommenders, without any linguistic analysis or automatic assessment.

2. LEMI – a corpus based literacy support tool

While readability studies in English are not new, research for other languages such as Romanian is scarce. However, studies such as Botarleanu et al. 2023 have taken the first steps towards developing methods in order to measure the complexity of words within texts across different languages (p. 2). This is where the LEMI project (*Reading for Me. Science for Children*, coordinated by the West University of Timisoara, Romania) will fill in the research gap and become "the first Romanian tool that uses computational linguistics methods to assess school children's literature complexity and readability" (Chitez et al., 2023, 2), in order to achieve a "correct pairing of textbook complexity and student grade level" (Paraschiv et al. 2023, 52). LEMI's technical profile is a SaaS that uses Machine Learning (ML) and Natural Language Processing (NLP) methods to automatically assess young children (age 7-11) literature's complexity, readability, and age adaptability. Moreover, it will offer an online platform with a digital repository of school reading, useful for children, teachers, NGOs, and other stakeholders, who will have immediate access to any new forms of written culture. Not only will the available literary corpus be linguistically assessed, but LEMI will provide an automated evaluation of the reading level for texts selected/uploaded by users.

3. Testing readability formulas for Romanian

The main challenge in this context is to develop the criteria for the automated assessment. In order to test the available readability formulas, the research team selected texts already included in school curriculum (several texts present in textbooks designed for different classes). For instance, an excerpt from a classical author such as Ion Creangă (adapted from *Amintiri din copilărie*) was included in a 3rd grade textbook (Figure 3).

Amintiri din copilărie

– fragment –

după Ion Creangă

Vocabular

iarmaroc: târg;
a o tuli: a fugi;
tolănit: întins într-o poziție confortabilă;
prund: mal, țărm, teren acoperit cu pietriș;
tupiluș: pe furiș;
păpușoi: loc plantat cu porumb.

Într-o zi, pe aproape de Sânt-Ilie, se îngrămădise, ca mai totdeauna, o mulțime de treburi pe capul mamei. Și mă scoală mama atunci, mai dimineață decât alte dăți, și-mi zice cu toată inima:

— Nică, dragul mamei! vezi că tată-tău e dus la coasă, căci se scutură ovăzul, și eu nu-mi văd capul de treburi! Tu mai lasă drumurile și stai lângă mămuca, de leagănă copilul; c-apoi și eu ți-oi lua de la **iarmaroc** o pălăriuță cu pană și o curălușă!

— Bine, mamă! Dar în gândul meu numai eu știam.

Când auzeam de legănat copilul, nu știu cum îmi venea; căci tocmai pe mine căzuse păcatul, să fiu mai mare între frați. Însă ce era să faci, când te roagă mama?

Dar în ziua aceea, în care mă rugase ea, era un senin pe cer și așa de frumos și de cald afară, că-ți venea să te scalzi pe uscat, ca găinile. Văzând eu o vreme ca asta, am **tulit**-o la baltă.

Figure 3. Textbook excerpt, 3rd grade, Aramis publishing house. https://manuale.edu.ro/manuale

Readability formulas that are widely used for English and available on free online platforms delivered inconclusive results, because the same text is classified in multiple 'grades' or 'classes' (see Table 1). All formulas focus on linguistic factors, such as word length and sentence length: The Flesch Reading Ease formula (Flesch, 1948) and the Gunning Fog Index (Gunning, 1952) use sentence length; the Simple Measure of Gobbledygook (McLaughlin, 1969) takes into account the number of syllables per word and number of polysyllabic words; while the Coleman-Liau Index (Coleman & Liau, 1975) and the Automated Readability Index (Smith & Senter, 1967) quantify characters instead of syllables per word. Using the metrics available online not only confuses the reader in terms of text classification, but the calculations are not appropriate for Romanian, as the manual scores and calculations show in Table 1.

Table 1. Online available automated readability calculations.

Formula	Manual score	Online score provided by readabilityformulas.com
Flesch Reading Ease (FRE)	50.81244	72.8
Gunning Fog Index (GFI)	15.75385	9.7
Coleman-Liau Index (CLI)	7.503671	1
Simple Measure of Gobbledygook (SMOG)	8.196152	6.5
Automated Readability Index (ARI)	7.705217	2.5

When manually calculating the number of words, syllables, letters, and sentences in the excerpt (see Table 2), we concluded that the main problem with the available readability formulas is quantifying the number of syllables in a low-resource language like Romanian. Hence, the main challenge in creating the support tool for Romanian will be to integrate an appropriate syllable separator and counter into an existing readability formula.

Table 2. Manual calculations of syllables and letters.

Sentence (separator . ;)	Words	Syllables	Letters (without hyphens)
1	18	35	82
2	15	29	65
3	3	6	15
4	18	27	68
5	11	22	51
6	15	26	55

7	2	4	8
8	7	10	26
9	10	17	43
10	12	20	51
11	9	13	30
12	31	48	107
13	10	16	36
TOTAL	161	273	637
AVERAGE	12.38461538	21	49
Words with 3 or more syllables	27		
syllables/words	1.695652174		
letters/words	3.956521739		

4. Discussion and conclusions

The lack of consistency provided by existing readability formulae requires researchers to develop specific tools for the Romanian language. Which of the existing formulae is the most appropriate? Are any of them suitable for Romanian? What would a specific formula for Romanian look like? How do we define the reporting scale/scheme? – these are key questions for designing LEMI's main automated formulae. Since the currently available options are not sufficient, the development of a new readability formula for Romanian will take into account 'traditional' criteria, such as text, sentence and word length, as well as alternative, but necessary measures, such as "word maturity", "age of exposure word lists" (see Botarleanu 2023). The latter aspects will focus more on comprehension and text complexity in terms of semantics, while the traditional metrics will address the morphological and, more generally, the grammatical dimension of a text. Moreover, classifying children's literature with the help of BERT-based models (Bidirectional Encoder Representation from Transformers) (Paraschiv et al. 2023) will enhance the innovative character of LEMI, by providing the testing and scientific validation of the first method of assessing the complexity of school texts in Romanian. The computational validation of a new formula within the ReaderBench framework will be complemented by a process of 'user validation', where students will have access to a beta version of the LEMI platform and will be able to individually rate the texts from the compiled corpus. This hybrid, multi-stage validation will refine the development of an exclusive concept of technological transfer from the scientific area to the educational one, in the field of early school reading.

Acknowledgements

PROIECT CO-FINANȚAT DE: Administrația Fondului Cultural Național

This work was supported by a grant of the Administration of the National Cultural Fund (AFCN) of the Romanian Ministry of Culture, in the framework of the programme *Promotion of Written Culture*, session I/2023, for the project LEMI (*Lectură pentru mine. Știința în slujba copiilor* – LEMI/ *Reading for Me. Science for Children*; January-November 2023), contract no. P0299/10.02.2023. The project was awarded to the West University of Timisoara for the proposal submitted by the project coordinator, CS II Dr. Habil. Madalina Chitez.

References

Botarleanu, R. M., Dascalu, M., Watanabe, M., Crossley, S. A., & McNamara, D. S. (2022). Age of Exposure 2.0: Estimating word complexity using iterative models of word embeddings. *Behavior research methods*, *54*(6), 3015–3042. https://doi.org/10.3758/s13428-022-01797-5.

Chitez, M., Rogobete, R., & Oravitan, A. (2023, June). Designing LEMI: the Romanian language tool that makes kids love reading. In *Conference Proceedings. The Future of Education 2023*. https://conference.pixel-online.net/files/foe/ed0013/FP/3177-PRI6037-FP-FOE13.pdf.

Coleman, M. and Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*(2), 283–284. https://doi.org/10.1037/h0076540.

DuBay, W. H. (2004). *The principles of readability*. *Online Submission*. Available at: https://files.eric.ed.gov/fulltext/ED490073.pdf.

European Commission, Directorate-General for Education, Youth, Sport and Culture (2022). *Education and training monitor 2022: Romania, Publications Office of the European Union*. Retrieved October 10, 2023, from https://data.europa.eu/doi/10.2766/310121.

Eurostat (2022). *Early leavers from education and training. Statistics Explained*. Retrieved October 10, 2023, from https://ec.europa.eu/eurostat/statistics-explained/SEPDF/cache/1150.pdf.

Flesch, R. (1948). A new readability yardstick, *Journal of Applied Psychology*, 32(3), 221–233. https://doi.org/10.1037/h0057532.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.

Ho, E. S. C., & Lau, K. (2018). Reading engagement and reading literacy performance: effective policy and practices at home and in school. *Journal of Research in Reading*, *41*(4), 657–679. https://doi.org/10.1111/1467-9817.12246.

Klare, G. R. (1963). *The measurement of readability*. Ames, Iowa: Iowa State University Press.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, *12*(8), 639–646.

OECD (2009). *PISA 2009 reading framework*. Paris: OECD Publications.

Paraschiv, A., Dascalu, M., & Solnyshkina M. I. (2023). Classification of Russian Textbooks by Grade Level and Topic Using Readerbench. *Научный результат. Вопросы теоретической и прикладной лингвистики (Research Result. Theoretical and Applied Linguistics)*, *9*(1), 73–86. https://doi.org/10.18413/2313-8912-2023-9-1-0-4.

PISA IDE. (n.d.) *Averages for age 15 years PISA reading scale: overall reading, by All students [TOTAL] and jurisdiction: 2018 and 2015*. Retrieved July 31, 2023, from https://pisadataexplorer.oecd.org/ide/idepisa/report.aspx?p=1-RMS-1-20183,20153-PVREAD-TOTAL-IN2,IN3-MN_MN-Y_J-0-0-37&Lang=1033.

Smith, E. A., & Senter, R. J. (1967). *Automated readability index*. Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command. Wright-Patterson Air Force Base, Ohio.

Ziming, L. (2005) Reading behavior in the digital environment. *Journal of Documentation*, *61*(6), 700–712. http://dx.doi.org/10.1108/00220410510632040.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16991

A machine-learning approach to Czech readability

Peter Williams[a] and Robert Reynolds[b]
[a]Department of Linguistics, Brigham Young University, peterjwms@gmail.com and [b]Office of Digital Humanities, Brigham Young University, robert_reynolds@byu.edu

How to cite: Williams, P.; Reynolds, R. (2023). A machine-learning approach to Czech readability. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16991

Abstract

We present a new corpus of Czech texts labeled for second-language readability, and show results of experiments to train machine-learning classifiers to automatically label new texts according to reading level. We report results comparing the performance of traditional machine-learning models (including Random Forest, XGBoost, Linear Discriminant Analysis, and XGBoost Random Forest) and a neural network (XLM-RoBERTa). The results of our research can be implemented in tools to support learning Czech, a less commonly taught language. We extract 46 linguistic features in various categories for use with traditional machine-learning algorithms. We train models on these features and evaluate their performance with recursive feature elimination to determine how informative each feature is for each model. We then compare those results to those of a transformer trained for the same task on the same corpus. XGBoost achieves the highest accuracy at 0.81, suggesting that these traditional models can still perform as well as, or better, than newer models on this task. Notably, the transformer has the lowest mean F1 at 0.74.[1]

Keywords: *readability, machine learning, Czech, transformer, corpus.*

1. Introduction

Traditional approaches to readability assessment have used formulae based on average word length and average sentence length. Although these features are useful, they represent an impoverished view of readability. Recent years have seen a marked increase in interest in research surrounding machine-learning approaches to second-language readability, with many studies focused on commonly taught languages, such as English (Vajjala & Meurers, 2012; Xu et al., 2015; Xia et al., 2016; Vajjala & Lučić, 2018), German (Hancke et al., 2012), Russian (Reynolds, 2016), French (Lee & Vajjala, 2022), Italian (Azpiazu & Pera, 2019), and Spanish (Vásquez-Rodríguez et al., 2022). In addition to studying the readability of individual languages, some researchers have worked to identify language-agnostic properties that can be used for multilingual readability classification (Azpiazu & Pera, 2019).

As research in this field continues, it expands both in methods and languages. Some research has been done recently to adapt classic readability metrics to Czech, including the Flesch Reading Ease, Flesch-Kincaid Grade Level, Coleman-Liau Index, and the Automatic Readability Index (Bendová & Cinková, 2021). In addition to traditional machine-learning models, neural networks have also been applied to the readability assessment problem in English and Slovenian, and recently achieved similar levels of success as traditional machine-

[1] Code available at https://github.com/peterjwms/czech-readability.

learning approaches (Deutsch et al., 2020; Martinc et al., 2021; Lee et al., 2021). In order to push the limits of both traditional machine-learning approaches with linguistic features and neural approaches, a combination of linguistic features with a neural network has been used with similar results (Lee et al., 2021).

We contribute to these threads of research in two ways. First, our study is the first to apply machine-learning methods to readability in Czech, a less commonly taught language which is typologically distinct from most languages already studied in the readability literature. Second, a labeled second-language readability corpus in any new language represents a contribution toward efforts to build multilingual readability models.

Based on this corpus, we train a number of machine-learning classifiers to label texts according to readability level. On the one hand, we are interested in maximizing classification accuracy, but we also use the models to help determine which document features are most informative for this task. To this end, we use both neural and traditional machine-learning approaches. For each document, we use natural language processing to extract features in the following categories: Lexical Variability, Lexical Familiarity, Morphology, and Syntax.

First, we use these features to create a text classification model using traditional machine-learning algorithms. Using the coefficients/importances of these models, we evaluate our features, identifying which features are most informative for this task. Second, we compare these results against those of a transformer. As the first study of Czech L2 readability using machine learning, this study lays the groundwork for both theoretical and technical research supporting the learning of a less commonly taught language.

2. Method

2.1. Corpus collection

Our corpus comes from two sources. First, we collected graded readers with parallel translations of Czech and other languages labeled with CEFR readability level of either A1-A2 or B1-B2 by the publisher. Because these labels span sublevels 1-2, we simplify the labels to 'A' and 'B'. Each of these books came from the Czech publisher Edika[2], part of Albatros Media. Each chapter was treated as its own document in order to normalize the lengths of each document and increase the number of documents in the training corpus. Second, we filled the gap of C-level text using Czech news articles from the Czech Text Document Corpus 2.0 (CTDC) (Kral & Lenc, 2017). The CTDC includes newspaper articles from the Czech News Agency labeled topically at the document level. We assume that all documents from CTDC are at the C reading level. We randomly selected a subsection of the CTDC to have a mostly balanced number of texts for each level, with a variety of topics, themes, and styles present in each level. Our corpus has 229 A-level documents, 209 B-level documents, and 230 C-level documents, with a total of 426,834 tokens. The number and level of classes were determined by their availability in our dataset.

2.2. Feature extraction

In total, we extracted 46 features split into several categories: lexical variability, lexical familiarity, morphology, and syntax. Lexical variability includes features like type-token ratio. Lexical familiarity includes word and lemma frequencies and ranks, obtained from the corpus SYN2015 from the Czech National Corpus[3] (Český národní korpus, 2016). Morphology includes the average number of certain parts of speech per sentence, and part-of-speech to token ratios. Syntax includes mean and max dependency lengths for various parts of speech. These features were extracted for each chapter primarily using Stanza (Qi et al., 2020).

Additionally, we used several classic measures of readability, with formulas adapted to Czech by Bendová and Cinková (2021). These formulas include adjustments to the coefficients of the Flesch Reading Ease, Flesch-

[2] https://www.edika.cz/

[3] https://wiki.korpus.cz/doku.php/seznamy:srovnavaci_seznamy

Kincaid Grade Level, Automatic Readability Index, and the Coleman-Liau Index. The function to count Czech syllables for the purpose of these readability metrics was adapted from Bendová[4] (2021).

2.3. Choice of models

For the machine-learning models, we used Linear Discriminant Analysis (LDA) and Random Forest (RF) models from scikit-learn (Pedregosa et al., 2011), and XGBoost (XGB) and XGBoost Random Forest (XGBRF) models from XGBoost (Chen & Guestrin, 2016). For the transformer, we used XLM-RoBERTa (base size) from HuggingFace as the pre-trained model on which we fine-tuned (Conneau et al., 2019).

2.4. Folds

The chapters in each book are likely to share similar vocabulary, grammar, and style, so we structured our cross-validation folds to avoid having chapters from the same book in both the training and validation. We grouped the texts in our corpus according to the book that they originally belonged to, such that all the texts belonging to the B-level 'Jana Eyrova' are part of group 2, and a random selection of articles from the CTDC are group 33, and so on. Using these groups, we manually created folds for cross-validation that combined one group each of A, B, and C texts. One fold was kept out as a test set. The rest of the data was used for cross-validation. In this way, we were able to control for specific variations in style and content that would affect reading level. This method of cross-validation was applied to both the traditional machine-learning models as well as the transformer.

We also performed recursive feature elimination on each model, and then trained the models again by adding the next most informative feature for that model each loop. Each model was trained with the same folds and cross-validation method as before.

3. Results

Table 1. Performance metrics on cross-validation

Model	Accuracy	Mean precision	Mean recall	Mean F1
LDA	0.77	0.91	0.77	0.80
RF	0.80	0.90	0.80	0.83
XGB	**0.83**	0.91	0.83	**0.85**
XGBRF	0.78	0.89	0.78	0.80
XLM-RoBERTa	0.75	0.78	0.74	0.74

The results from training each model are shown in Table 1. F1 is the harmonic mean of precision and recall. These standard metrics are calculated as a weighted average of the performance of each fold on the validation set, and then a macro average was taken of those to find an overall score for each model. Across all metrics, XGB achieves the best performance, with a mean accuracy of 0.83 and mean F1 of 0.85. XLM-RoBERTa performs slightly worse on all metrics than all of the traditional machine-learning models. Notably, each traditional model has a much higher precision than recall, which means that the models are returning very few false positives, but are all missing some of the true positives.

[4] https://github.com/vanickovak/ReadabilityFormula/tree/main

When performing recursive feature elimination on the results of each model, we see that each model reaches similar performance, but using different combinations and numbers of features to reach peak performance, as is visible in Figure 1. RF reaches peak performance with only 12 features with peak mean F1 score of 0.84. XGBRF is close behind with a peak mean F1 of 0.82 at 20 features. With similarly few features, LDA and XGB have F1 scores within 0.05, and then both peak much later with 31 and 46 features at 0.82 and 0.84.

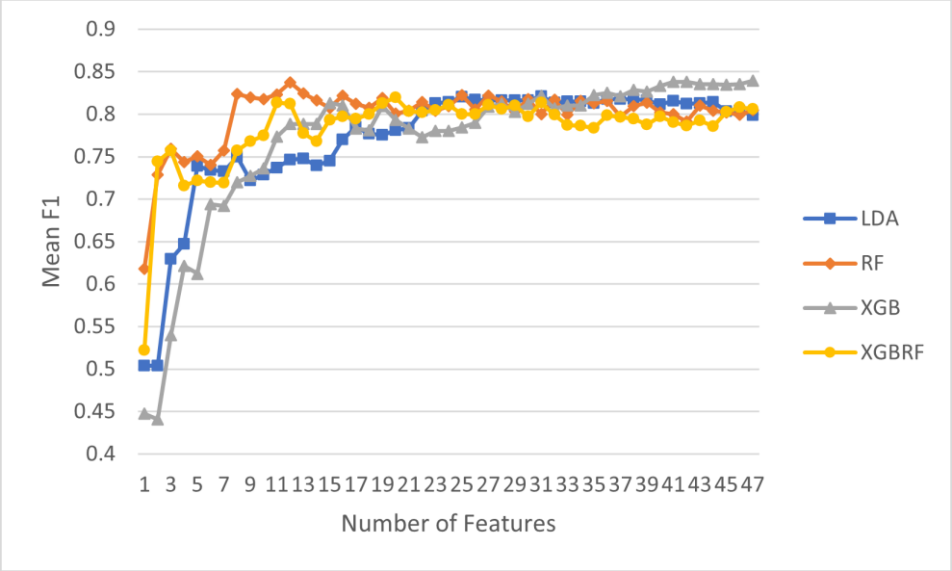


Figure 1. Model performance by number of features

We see that of the ten most informative features for XGB, RF, and XGBRF, many are the same for all three, including: mean lemma frequency, pronoun-token ratio, mean number of nouns per sentence, mean number of indicative verbs per sentence, mean number of adpositions per sentence, mean dependency length, Automatic Readability Index, and Coleman-Liau Index.

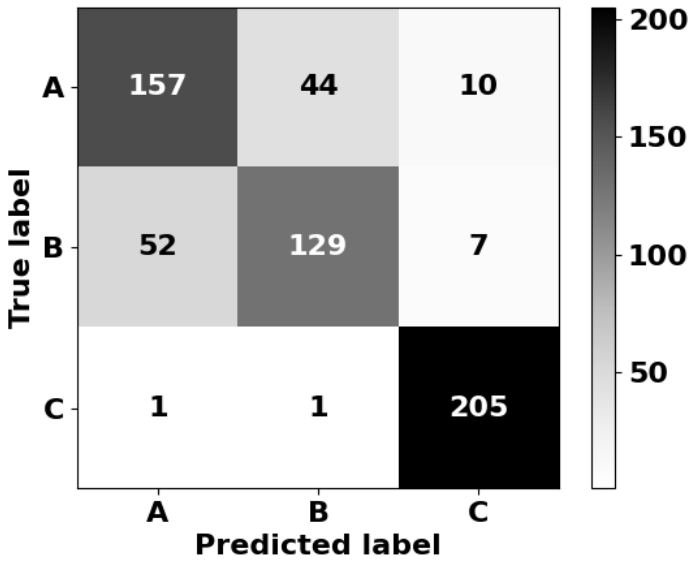


Figure 2. XGBoost cross-validation confusion matrix

4. Discussion

Our results suggest that for this problem, XGB is the best-performing model, especially considering the higher mean F1 score and consistently high performance in standard training, and when considering feature importances. RF performs comparably when considering feature importances and training the model to use as few resources as possible. The two random forest architectures perform better on low resources than our other models, as expected, although XGB outperforms them both with more data. The consistently lower results from XLM-RoBERTa suggest that the traditional models with handcrafted features are currently better-suited for this task in Czech. We could potentially improve performance by fine-tuning a different multilingual or Czech-specific model, or by augmenting the transformer architecture with linguistic features.

The confusion matrix in Figure 2 exhibits a pattern seen in all of our models: the majority of the errors confuse A and B texts. To ensure that these errors were not due to a bad assumption that all of the news documents are level C, we trained binary classifiers with only documents at levels A and B. Results were similar, which suggests that either some of the A- and B-level texts are mislabeled, or that we have failed to identify features that distinguish between these levels. Additionally, when certain books are used as the validation set, all of the models perform much worse, suggesting that these books in particular might be poorly labeled, or otherwise unique enough in their features that our models were not able to account for them. These errors might have been mitigated by collection of a larger or more varied corpus, or by confirming that each text was labeled accurately.

From our recursive feature elimination, we notice that the most informative features for XGB, XGBRF, and RF span all of our categories of features, including readability metrics, suggesting that none of the categories is considerably less informative. LDA, on the other hand, overlaps with the other three only in its use of pronoun-token ratio, and otherwise heavily relies on part-of-speech to token ratios and word length metrics.

5. Conclusions

We trained traditional and neural machine-learning models for labeling Czech documents according to three readability levels. XGBoost achieved an accuracy of 0.83. Further work can focus on increasing the size and accuracy of the labeled corpus, and testing other models, including a hybrid architecture that combines linguistic features with transformers, as demonstrated on English by Lee et al. (2021).

Acknowledgements

We would like to thank the BYU College of Humanities for their assistance in funding the corpus collection and travel to the conference.

References

Azpiazu, I. M., & Pera, M. S. (2019). Multiattentive recurrent neural network architecture for multilingual readability assessment. *Transactions of the Association for Computational Linguistics*, *7*, 421-436.

Bendová, K. (2021). Using a parallel corpus to adapt the Flesch Reading Ease formula to Czech. *Jazykovedný časopis, 72*(2), 477-487.

Bendová, K., & Cinková, S. (2021, August). Adaptation of classic readability metrics to Czech. In *International Conference on Text, Speech, and Dialogue* (pp. 159-171). Cham: Springer International Publishing.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Český národní korpus: *Srovnávací frekvenční seznamy*. Ústav Českého národního korpusu FF UK, Praha 2016. Dostupné z WWW: http://www.korpus.cz

Deutsch, T., Jasbi, M., & Shieber, S. (2020). Linguistic features for readability assessment. *arXiv preprint arXiv:2006.00377*.

Hancke, J., Vajjala, S., & Meurers, D. (2012, December). Readability classification for German using lexical, syntactic, and morphological features. In *Proceedings of COLING 2012* (pp. 1063-1080).

Lee, B. W., Jang, Y. S., & Lee, J. H. J. (2021). Pushing on text readability assessment: A transformer meets handcrafted linguistic features. *arXiv preprint arXiv:2109.12258*.

Lee, J., & Vajjala, S. (2022). A neural pairwise ranking model for readability assessment. *arXiv preprint arXiv:2203.07450*.

Král, P., & Lenc, L. (2017). Czech text document corpus v 2.0. *arXiv preprint arXiv:1710.02365*.

Martinc, M., Pollak, S., & Robnik-Šikonja, M. (2021). Supervised and unsupervised neural approaches to text readability. *Computational Linguistics, 47*(1), 141-179.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Reynolds, R. (2016, June). Insights from Russian second language readability classification: complexity-dependent training requirements, and feature evaluation of multiple categories. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 289-300).

Vajjala, S., & Lučić, I. (2018, June). OneStopEnglish corpus: A new corpus for automatic readability assessment and text simplification. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications* (pp. 297-304).

Vajjala, S., & Meurers, D. (2012, June). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the seventh workshop on building educational applications using NLP* (pp. 163-173).

Vásquez-Rodríguez, L., Cuenca-Jiménez, P. M., Morales-Esquivel, S., & Alva-Manchego, F. (2022, December). A Benchmark for Neural Readability Assessment of Texts in Spanish. In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)* (pp. 188-198).

Xia, M., Kochmar, E., & Briscoe, T. (2019). Text readability assessment for second language learners. *arXiv preprint arXiv:1906.07580*.

Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics, 3*, 283-297.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16961

Co-creating CALL content - does it work? Goldilocks compromise or Cruella chaos?

Monica Ward[a], Elaine Uí Dhonnchadha[b], Jennifer McGarry[c] and Liang Xu[d]
[a]Dublin City University, monica.ward@dcu.ie; [b]Trinity College Dublin, uidhonne@tcd.ie; [c]Dublin City University, jennifer.mcgarry@dcu.ie and [d]Dublin City University, liang.xu6@mail.dcu.ie

How to cite: Ward, M.; Uí Dhonnchadha, E.; McGarry J.; and Xu, L. (2023). Co-creating CALL content - does it work? Goldilocks compromise or Cruella chaos? In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16961

Abstract

Developing Computer-Assisted Language Learning (CALL) resources is a complex task that requires a multidisciplinary team. Each team member brings their unique perspective – the teacher focuses on pedagogy, the developer on software, the Natural Language Processing (NLP) expert on NLP, the game developer on gaming, and the CALL expert strives to maintain a balance between all of these elements. To create effective CALL resources, collaboration and co-creation among team members is crucial. This paper delves into the intricacies of CALL resource development through the lens of a digital, game-based CALL resource for Irish, undertaken by a team comprising a teacher, a game developer, an NLP specialist, and a CALL researcher. It explores the challenges, opportunities, and the delicate balancing act required during development, to achieve a 'Goldilocks compromise' rather than 'Cruella de Vil chaos'.

Keywords: *CALL, less commonly taught languages, educational games, co-creation.*

1. Introduction

Computer-Assisted Language Learning (CALL) development is a multidisciplinary endeavour and ideally involves a team with different skill sets (Ward, 2015a). However, even for the most commonly taught languages, this is often not the case, and for Less Commonly Taught Languages (LCTLs) it can be rare. Usually, CALL researchers for LCTLs have to use whatever resources are available and have to try to repurpose existing resources for their LCTL.

CALL resource development can also be frustrating for individual team members. Each member is usually most focussed on getting their part 'right', and having to make compromises with the other team members can sometimes require patience. The teacher wants to focus on pedagogy, the Natural Language Processing (NLP) specialist wants to focus on getting the NLP aspects right, the game-designer wants to make sure that the game element is kept to the fore, and the CALL researcher wants to ensure that learnings from CALL research underpin what is developed.

This paper looks at the development of Cipher, a digital, game-based CALL resource for Irish (Ward et al., 2022). Current teaching is textbook-based and students have very limited access to digital resources. In order to strengthen the alignment of the Cipher game with the classroom curriculum, the original team of a digital game developer, an

NLP researcher, and a CALL researcher was strengthened by the addition of a primary school teacher. The end goal for the team was the development of a game for learners of Irish that would be an attractive game, while at the same time being aligned with the school curriculum. The game texts are based on fairy stories and myths and the challenge is to weave the desired pedagogical components into the narrative in an unobtrusive manner, with emphasis on maintaining the game's ambiance. We wish to avoid a 'chocolate-covered broccoli' situation (Hopkins, 2015), whereby it becomes a gamified language learning resource which duplicates the curriculum, rather than a CALL resource which complements the curriculum and enhances the learning experience.

This paper reports on the co-creation process used to ensure better alignment with the school curriculum. It looks at whether a 'Goldilocks compromise' was reached or whether 'Cruella de Vil chaos' ensued, and provides suggestions for other CALL researchers considering adopting a similar approach.

2. Background

2.1. Software development and CALL

In the early days of computing, software development was very linear. The Waterfall model (e.g. Sommerville, 2004) was the dominant model. Needs were analysed, the system was designed, developed, tested, delivered to the end-user, and finally entered the maintenance phase. While this seems like a logical approach, there was a long time between the initial analysis phase and the final delivery of software to the client, often with little interaction between the development team and the final end-user. More recently, the agile approach has come to the fore where development occurs in iterative and incremental cycles, with frequent collaboration between teams and continuous feedback from end-users. Agile methodologies prioritise user satisfaction, teamwork, and the ability to respond effectively to evolving needs (Highsmith, 2002). This approach works well in situations where there are unknown elements or novel aspects for the development team. In the context of CALL, the development of resources is non-linear, particularly in the context of LCTLs. The development team may be ad-hoc, and the team members may have limited availability as they may have to share their time between several projects. It may be difficult to arrange team meetings with all of the required members present, and it might not be possible to carry out tasks in the preferred order. One benefit of the agile approach is that there is continual interaction between the developers and the client, so there is scope to clear up misunderstandings as the development progresses.

2.2. Co-creation

Co-creation implies the creation of something between different parties. In education, particularly at higher/third level, it refers to co-creation between educators and students in a student partnership model (Bovill, 2020). In the field of CALL, there are two types of potential end-user - educators/teachers and learners. The teachers know what they want to teach but may not know how to develop it. The developers know how to develop a resource but may not be familiar with the pedagogical requirements. An NLP specialist may not be familiar with CALL. Learners may not have the experience or maturity to understand the language learning process. In a co-creation approach, it is accepted that no one person knows everything needed to develop the desired system, but together they contribute their expertise to enhance the development of CALL resources, emphasising learner needs and the deployment context (Ward, 2015b).

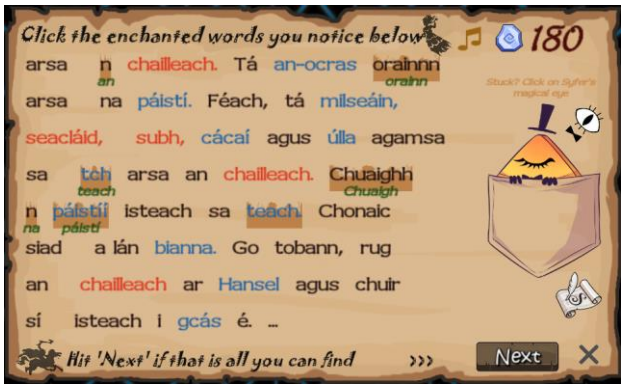
3. Method

Cipher was developed for English and then adapted and modified for Irish, with a particular focus on primary school learners (Ward et al., 2022). The first version of Cipher (Xu et al., 2022a) was relatively successful and students were able to use it in their own classroom. However, there was a desire to make sure that the content was pedagogically focussed and aligned with the school curriculum, as previous research indicates that this is important for the uptake of CALL resources in the school context (Ward, 2007). The Cipher team consisted of a digital, game-based language learning developer interested in learning motivation and cultural heritage, an NLP and Irish language expert, a primary school teacher, and a CALL researcher with an interest in resources for LCTLs. The

game developer did not speak Irish and had not attended school in Ireland. The other team members had no game development expertise. The Irish language expert was not familiar with Irish pedagogy while the teacher was not familiar with CALL resources for Irish. However, all members of the team wanted to develop a useful and usable CALL resource for Irish by adapting and aligning Cipher with the Irish primary school curriculum.

There were four iterative steps in the development process. Firstly, the game developer worked closely with the NLP and language specialist and CALL researcher to develop the initial version of each of the stories to be used in the Cipher game. The texts were chosen with the target language learners, primary school children aged 8-12, in mind. Secondly, the teacher reviewed the texts and suggested vocabulary words that could be added to the story, so that it would align with what was being taught in the classroom. Thirdly, there were discussions between the team members to strike the right balance between curriculum alignment and maintaining the game ambiance. In step 4, the final texts were agreed and added to the Cipher game.

Figure 1 shows part of an initial version of the Hansel and Gretel story in the game. As the story deals with food, the teacher suggested that more food words could be added to the story. While the addition of some words make sense, not all the words were added as this would change the focus of Cipher from a game to a vocabulary learning task. It would distort the story and dilute the game aspect of the CALL resource. Figure 2 shows part of the amended story in the game with the addition of some of the suggested words.



Translation:

"..., said the witch.

We are very hungry, said the children.

Look, I have sweets, chocolate, jam, cakes, and apples in the house, said the witch.

The children went into the house.

They saw a lot of foods. Suddenly the witch grabbed Hansel and put him in a cage …"

Figure 1. Initial version of the Hansel and Gretel story in the game along with translation

Translation:

"The witch came out.

Are you hungry? said the witch.

We are very hungry, said the children.

Look, I have sweets, chocolate, jam, cakes, crisps, and apples in the house, said the witch.

The children went into the house.

They saw a lot of foods - popcorn, pancakes ..."

Figure 2. Amended version of the Hansel and Gretel story in the game along with translation

4. Results and discussion

The Cipher game was tested in three primary school classrooms in Ireland. Cipher was installed on tablets and each student could play the game individually. Although there were issues with Wi-Fi access in one of the classrooms, the students enjoyed the game and the teachers reported that it was good for their students. Following a two-month period of weekly engagement with the game, students were invited to participate in a survey. We considered responses rated as 4 or 5 on a 5-point Likert scale as indicative of a positive assessment, aligning with

the criteria set forth in Xu et al. (2022b). Analysis of the responses indicated that the students reported high satisfaction ratings, particularly when comparing learning Irish through the game compared to usual classroom teaching (66%). See Table 1 for details.

Table 1: Distribution of ratings provided by the participants.

Question	Satisfaction (n = 165) positive (percentage)
Did you like playing the game?	53%
Do you like the images in the story?	51%
What do you think about learning Irish through the game?	39%
How would you compare learning Irish through the game to normal classroom teaching?	66%
How do you feel about learning Irish after playing the game?	38%
Do you think you learned anything while playing the game?	34%
Would you recommend this game to your friends?	39%

We also collected verbal and written feedback from the teachers in the classrooms. Comments from teachers included, "*it is very motivating for the children*," and they also had suggestions for improvement ("*I would need a few sample questions to get better before completing the game*!").

The co-creation aspect of the development was interesting and challenging. The teacher was an expert in her classroom and was familiar with teaching Irish in an English-medium primary school. She knew the curriculum and the level of students' knowledge of the language. The game developer stayed focussed on the game element, while being guided by the other team members in terms of Irish language content. The Irish language expert tended to overestimate the Irish language ability of the English-medium primary school students, while the CALL researcher was concerned with ensuring that the game worked, that it was pitched at the right level for the students, and was beneficial for their learning.

One clear example of the co-creation process in action was to do with the development of the content of the stories. The teacher was keen on ensuring that there were a good number of vocabulary words in each story. The game developer was concerned that this would detract from the game element. The NLP and Irish speaker was more concerned with maintaining the narrative aspect of the story. At each step of the development process, there were collegial discussions between each of the team members with each person outlining their point of view. Each member recognised that the other members had a different perspective and that it was important to take these into consideration and that compromise was necessary.

From the inception of the collaborative undertaking, it became evident that every team member would make distinct contributions to the CALL development process. The game designer's understanding of game dynamics and commitment to preserving the game's atmosphere was instrumental in shaping the project. The teacher had expertise in the school curriculum, and understood the limited time dedicated to Irish language instruction within the school timetable, and the essentiality of providing relevant resources to students. The Irish language specialist was instrumental in maintaining a consistent standard of Irish throughout the project's progression. The CALL researcher recognised that certain compromises would be necessary to achieve the intended outcome. Despite not precisely achieving the 'Goldilocks zone', a harmonious collaborative effort prevailed, steering clear of the disorderly tendencies and 'Cruella de Vil chaos'.

5. Conclusions

When developing CALL resources, it is important to have a multidisciplinary team. The team should adopt a co-creation approach, where each member is valued for their individual contributions and perspectives. CALL is a multi-disciplinary enterprise. No one person has all the knowledge required to develop a CALL resource, but with a co-creation perspective, where mutual respect is paramount and a shared vision reigns, there is no doubt that the final product will add up to more than the sum of its parts.

Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No.18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Bovill, C. (2020). Co-creation in learning and teaching: The case for a whole-class approach in higher education. *Higher education*, *79*(6), 1023-1037. https://doi.org/10.1007/s10734-019-00453-w

Campanelli, A. S., & Parreiras, F. S. (2015). Agile methods tailoring–A systematic literature review. *Journal of Systems and Software*, *110*, 85-100. https://doi.org/10.1016/j.jss.2015.08.035

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. *Behaviour and Information Technology*, *22*(6), 397-409. https://doi.org/10.1080/01449290310001624329

Highsmith, J. A. (2002). *Agile software development ecosystems*. Boston: Addison-Wesley Professional.

Hopkins, I., & Roberts, D. (2015). 'Chocolate-covered Broccoli'? Games and the teaching of literature. *Changing English*, *22*(2), 222-236. https://doi.org/10.1080/1358684X.2015.1022508

Sommerville, I. (2004). *Software engineering*. (7th ed). London: Pearson Education Ltd.

Ward, M. (2007). The integration of CL resources in CALL for Irish in the primary school context (Doctoral dissertation, Dublin City University).

Ward, M. (2015a). CALL and less commonly taught languages: challenges and opportunities. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy (pp. 549-552). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000391

Ward, M. (2015b). Factors in sustainable CALL. In A. Gimeno, M. Levy, F. Blin, & D. Barr (Eds) *WorldCALL: Sustainability and Computer-Assisted Language Learning* (pp. 132-151). London: Bloomsbury Academic.

Ward, M., Xu, L., & Uí Dhonnchadha, E. (2022). Game based language learning for Irish: noticing errors while playing. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 208–213). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1488

Xu, L., Uí Dhonnchadha, E., & Ward, M. (2022a). Faoi gheasa: an adaptive game for Irish language learning. In S. Moeller, A. Anastasopoulos, A. Arppe, A. Chaudhary, A. Harrigan, J. Holden, J. Lachler, A.

Palmer, S. Rijhwani, & L. Schwartz, L (Eds) *Proceedings of the Fifth Workshop on the Use of* Computational *Methods in the Study of Endangered Languages* (pp. 133-138). https://doi.org/10.18653/v1/2022.computel-1.17

Xu, L., Uí Dhonnchadha, E., & Ward, M. (2022b). User experience study of "Cipher: Faoi Gheasa", a digital educational game for language learning and student engagement. In *Proceedings of the 2nd Workshop on Games Systems* (pp. 5-8). https://doi.org/10.1145/3534085.3534339

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16982

Didactization of L2 French variation in a complex context

Alena Barysevich[a]

[a]School of Languages and Literatures, University of Guelph, Canada, , abarysev@uoguelph.ca

How to cite: Barysevich, A. (2023). Didactization of L2 variation in a complex context. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16982

Abstract

Even though variation is an inherent element of language and its different aspects have been the subject of numerous studies, its didactization in Second Language classrooms remains, in general, marginal. The context of teaching variation in a French as a Second Language (FSL) classroom in Ontario (Canada) is even more complex. What kind of French to teach: European or Canadian? What is the place of local varieties of Canadian French? This paper investigates the sociolinguistic variation in the speech of first-year FSL students in a Canadian university who have learned their second language primarily in an educational context. A range of social factors that influence sociolinguistic competence are explored. Pedagogical implications of observed results are discussed. Designing didactic materials that value work on varieties of spoken French is emphasized. The role of CEFR-oriented professional development in FSL teaching is highlighted.

Keywords: *FSL, Canadian university, sociolinguistic competence, lexical variation, CEFR.*

1. Introduction

As a component of communicative language ability, sociolinguistic competence refers to the sociocultural factors of language use (Council of Europe, 2018). Geographical, stylistic, and social variation at the lexical and grammatical levels is among the various aspects of sociolinguistic competence.

Although the development of sociolinguistic competence is encouraged by educational (Council of Europe, 2018; CMEC, 2010) and sociolinguistic research (Bayley & Regan, 2004; Mougeon et al., 2010), it is often neglected over linguistic competence in FSL classrooms in Ontario (Rehner, 2010, 2011; Rehner & Mougeon, 2003). The integration of the lexical variation into Canadian FSL classrooms certainly poses some fundamental challenges, especially when French is in a minority context (e.g. Ontario) and the French-speaking population is very heterogeneous (Statistics Canada, 2021). For example, in Ontario, different varieties of French are in diglossic relationships with each other: (i) although spoken by a minority of population, European French is considered the norm; (ii) varieties of Laurentian French (Ontario and Quebec French) are strongly devalued (linguistic insecurity); and (iii) other French varieties (e.g. African French) are not recognized.

Several studies on the speech of FSL learners in Ontario pointed out that the language acquired in the classroom is different from the one spoken outside of the classroom (Mougeon & Rehner, 2015). More specifically, FSL learners in Ontario classrooms are overexposed to (super)formal forms but lack awareness of the authentic language spoken outside the formal setting. This trend is systematic in a French-minority setting among FSL learners where exposure to authentic language is limited (Mougeon et al., 2010). Furthermore, the study by Rehner et al. (2021) points out the relevance of teachers' training based on the principles of the *Common European Framework of Reference for Languages* (CEFR) in the development of sociolinguistic competence among FSL learners.

2. Method

The objective of our project is twofold:

(1) To evaluate the level of sociolinguistic competence among first-year university FSL learners in a Canadian university.
(2) To propose pedagogical resources and tools to foster students' sociolinguistic competence.

To accomplish the first objective, an online survey was administered to all first-year students in the FSL program at the University of Guelph (Ontario, Canada)[1]. Participants included 43 first-year students who took 24 weeks of FSL courses. All participants are English L1 users (native or with high level of proficiency). Participants' French proficiency level approximately equates to an average CEFR level of A2. 83% of participants are female and 17% are male between 17 and 22 years old.

The online questionnaire (30 questions) was created and distributed via Qualtrics. The survey included multiple choice and association questions, Likert items, and an open-ended section. This survey was designed to:

- identify students' sociolinguistic profiles (e.g. gender, age, mother tongue, languages spoken at home)
- investigate students' academic and linguistic backgrounds in French (e.g. participation in exchange programs, type of French program (Core, Immersion or Intensive French), level of Diplôme d'études en langue française/DELF)
- elicit students' auto-evaluation of their language skills (e.g. oral and written comprehension and production, level of comfort with French in different situations)
- obtain quantitative data about the students' lexical preferences (as a component of sociolinguistic competence).

We also carried out a documentary analysis of teaching resources used in first-year FSL courses. Finally, we conducted semi-structured interviews with nine instructors teaching first-year courses to understand their attitudes towards the place of sociolinguistic competence in the FSL curriculum.

3. Results

The majority of participants (80.4%) have English as the mother tongue but they constantly speak at home an additional language other than French (e.g. Italian, Spanish, Arabic, Punjabi, or Polish). For all except two students, French is their third language. The majority of participants (74%) use at least two languages at home, none speak French at home. 56% of participants completed either the French Immersion or Core French education system. 70.5% of participants have more than five years of experience in French (34% among them have more than 12 years). An overwhelming majority of participants expressed confidence with reading and writing tasks but much less confidence in oral comprehension and production tasks (analyses results of the Likert items) and reported recurrent difficulties with understanding French spoken outside of the classroom. Most participants expressed feeling comfortable speaking French in the classroom and with teachers, but felt intimidated speaking with native speakers of French, individually or in groups. Participants reported that their preferred activity outside of the classroom was watching movies in French and listening to French music[2]. Our participants are not keen about social media in French or taking part in Francophone communities. Less than 35% participated in exchange programs or traveled to Francophone places in or outside of North America.

For the questions about lexical choices (see Table 1), the results indicated a clear preference for more

[1] Due to adherence to the University of Guelph Research Ethics Boads policies, our project obtained ethics approval to engage with participants.

[2] In comments sections, students elaborated on types of music they listen to and movies they watch. None of the answers mentioned sources coming from Canadian French (all are exclusively from France).

(super)formal forms, with 65% of participants using exceptionally (super)formal variants and 35% using a combination of formal and less formal forms (but with a significant predominance of more formal ones). The results also show the insignificant use of terms characteristic of Canadian (Laurentian) French (e.g. *allô 'hi', jaser/commérer 'to chat', chocolatine 'chocolate croissant', chicaner 'to dispute, to fuss', char 'car', tantôt 'soon, later'*).

Table 1. Lexical Preference Among FSL Learners [Total – 43 speakers]. **In bold** – informal and vernacular forms.

	Number of Occurrences	%
bonjour 'hello, hi' *salut* ***allô***	29 11 3	67% 25.5% 7%
bavarder 'to talk, to chat' ***jaser*** ***commérer***	30 10 3	70% 23% 7%
travail 'a workplace' *emploi* ***boulot*** ***job***	24 16 1 2	56% 37% 2% 5%
auto 'a car' *voiture* *automobile* ***bagnole*** ***char***	13 30 0 0 0	30.2% 69.8% 0 0 0
pain au chocolat 'chocolate croissant' ***chocolatine***	38 5	88% 12%
rire 'to laught' ***rigoler***	39 4	90% 9%
patron 'manager, boss' *chef* ***boss***	13 23 7	30% 53% 16%
nous 'we' ***on***	15 28	35% 65%
alors 'so' *donc* ***so***	19 24 0	44% 56% 0
Allons-y! 'Let's go!' ***On y va!***	29 14	67% 33%

The crosstabulation of the data indicated that students who participated in exchange programs in Francophone places across Canada (e.g. Chicoutimi, Trois-Pistoles) are more familiar with lexical variability. Compared to students who completed Core French schools and showed no awareness of Canadian features, French Immersion

students seemed to be more aware of variability but prefer to imitate the instructors' speech and to use language acquired from manuals. Our literature review of the teaching materials used in FSL first-year courses shows that the language taught in the classroom is far from representing the linguistic reality outside of the classroom. This finding replicates Detey (2017) and Rehner & Mougeon (2003).

4. Discussion and Recommendations

Our study investigated the sociolinguistic competence of first-year FSL students in a Canadian university. One important finding was the observation of students' weak (or lack of) sociolinguistic competence. Although, with 5 to 12 years in FLS education, first-year university students have no basic awareness of existing variation across varieties of French language. Sadly, they are not even aware of basic characteristics of Canadian (Laurentian) French. The gap between students' acquired French in the classroom and the language spoken outside of the classroom is significant. As a result, FSL learners have low proficiency in oral comprehension and production and experience important difficulties in understanding French outside of the classroom. They feel highly frustrated and demotivated, with no interest to integrate in Francophones community outside of the formal setting. Our findings are aligned with studies pointing out insufficient educational input of the authentic language (Mougeon et al., 2010; Pöll, 2005).

A combination of factors may have contributed to the observed weak sociolinguistic competence. Firstly, as mentioned above, there is an insufficient input to the authentic language via educational materials. Our content analyses of the didactic materials used in the classroom determined a striking lack of exposure to the authentic use of Canadian French (e.g. no mention of lexical or grammatical variation, limited number of tasks exploring the use of Canadian French, and no explicit instruction of sociolinguistic competence). The analyzed educational materials suggest that students do not receive sociolinguistically-oriented FSL teaching. Secondly, favouring (super)formal variants can be attributed to the imitation of FSL instructors' speech, as many FSL instructors in Ontario have French as L2 or L3, and their speech is not always representative of Canadian French. Thirdly, many FSL instructors are not familiar with sociolinguistically-oriented FSL teaching promoted by CEFR. Also, limited exposure to authentic language outside of the classroom in a French-minority province, such as Ontario, represents a huge difficulty for the development of sociolinguistic competence among FSL students. Finally, negative attitudes (from students and instructors) towards varieties of Canadian French and the preference for idealized French spoken in Europe is an additional challenging factor in French-minority communities in Canada (Arnott, 2016; Arnott et al., 2019). In line with that, interesting findings emerged from our analysis of instructors' interviews: (i) general pro-normative attitude towards the European French norm rather than Canadian French; (ii) avoidance of less formal or vernacular forms; and (iii) the lack of CEFR and sociolinguistically oriented training (auto-perception).

Our results suggest that if the goal of FSL teaching is to form active participants of Francophone communities and to make our FSL learners capable of functioning in Canadian French communities, it is by no means achieved. We urge for a reconsideration of learning objectives and teaching practices in FSL programs for targeting the development of sociolinguistic skills. For example, to raise FSL learners' awareness of sociolinguistic aspects of French we call for more representation of Canadian French in educational materials (e.g. explicit and contextualized awareness of lexical variability via authentic documents and corpora). It could be done by using various free online authentic teaching materials with a high degree of variability targeting oral comprehension and production based on everyday Canadian (Laurentian) French (FrancoTolie; LISEO; PFC Enseignement du français; Corpus divers; Bazzo.tv; Radio-Canada; Tou.tv; Briser le code; Kebec).

To raise FSL learners' sociolinguistic awareness, real-life language tasks can take the form of brainstorming or research activities on lexical variation. Play activities can take many forms to teach a variety of sociolinguistic skills (e.g. board games, card games, and role-playing). One example of a play activity is the card game *Francophonies: Le grand jeu de toutes les langues françaises*, which introduces players to lexical variations across the French-speaking world. Short exchange programs in Canadian French communities (e.g. Explore) should be mandatory and systematic to increase interaction with the target language in order to develop positive

attitudes toward the target language and develop sociolinguistic competence. Finally, systematic extracurricular activities within Francophone communities should be an integral part of the FSL curriculum.

5. Conclusions

The results of our study show that the need for updating educational materials and pedagogical practices in FSL teaching in Ontario is undeniable to better reflect the Canadian reality. Our study emphasized the instrumental importance of the shift toward sociolinguistic-oriented language teaching and learning with the focus on authentic listening and speaking tasks and reevaluated assessment. We proposed a list of pedagogical resources for further developing sociolinguistic competence for first-year students at the post-secondary level. We also call for increased interaction with the target language during exchange programs so that learners are able to be part of Francophone community outside of the classroom.

The results from interviews with university instructors also showed the relevance of providing professional development for FSL instructors to interpret action-oriented pedagogy and sociolinguistic-driven teaching promoted by the CEFR. Finally, we take the opportunity to stress the undeniable importance of collaborative work between sociolinguists working on Canadian French and FSL educators in Canadian classrooms.

It is hoped that this paper might be used to reduce learning barriers FSL students encounter outside of classrooms. Although the study had a limited number of participants, its results replicated several studies on sociolinguistic competence among pre-university FSL students in Ontario, Canada. While this is promising, research investigating the factors driving sociolinguistic competence is a necessary step in further evaluating its efficacy.

Acknowledgment

I would like to thank my Graduate Research Assistant, Mackenzie Turner for taking part in this study.

References

Arnott, S. (2016). Second language education and micro-policy implementation in Canada: The meaning of pedagogical change. *Language Teaching Research*, *21*(2), 258-284. https://doi.org/10.1177/1362168815619953

Arnott, S., Masson, M., & Lapkin, S. (2019). Exploring trends in 21st century Canadian K-12 FSL research: A research synthesis. *Canadian Journal of Applied Linguistics*, *22*(1), 60-84. https://doi.org/10.7202/1060906ar

Bayley, R., & Regan, V. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics*, *8*(3), 323–338. https://doi.org/10.1111/j.1467-9841.2004.00263.x

Conseil des ministres de l'Éducation (Canada). (2010). L'exploitation du Cadre européen commun de référence pour les langues (CECR) dans le contexte canadien. http://cmec.ca/docs/CECR-contexte-canadien.pdf

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe Publishing.

Detey, S. (2017). La variation dans l'enseignement du français parlé en FLE : des recherches linguistiques sur la francophonie aux questionnements didactiques sur l'authenticité. In A.-C. Jeng, B. Montoneri & M.-J. Maitre (Eds.), *Échanges culturels aujourd'hui : langue et littérature* (pp. 93-114). New Taipei City: Tamkang University Press.

Mougeon, R., Nadasdi, T. & Rehner, K. (2010). *The sociolinguistic competence of immersion students*. Multilingual Matters.

Mougeon, R., & Rehner, K. (2015). The neglected topic of variation in teacher classroom speech: Investigation *je vais/je vas/m'as* in Ontario French-medium high schools. In R. Torres Cacoullos, N. Dion, & A. Lapierre (Eds.), *Linguistic variation: Confronting fact and theory* (pp. 165-177). New York, NY: Routledge.

Pöll, B. (2005). *Le français langue pluricentrique ? : Études sur la variation diatopique d'une langue standard*. Francfort: Peter Lang.

Rehner, K. (2010). The use/non-use of ne in the spoken French of university-level learners of French as a second language in the Canadian context. *Journal of French Language Studies, 20*(3), 289-311. https://doi.org/10.1017/S0959269510000025

Rehner, K. (2011). The sociolinguistic competence of former immersion students at the post-secondary level: the case of lexical variation. *International journal of bilingual education and bilingualism, 14*(3), 243-259. https://doi.org/10.1080/13670051003787566

Rehner, K., Lasan, I., Popovich, A., & Palta, Z. (2021). The Impact of CEFR-Related Professional Learning on Second-Language Teachers' Classroom Practice: The Case of French in Canada. *Canadian Journal of Applied Linguistics, 24*(1), 26–53. https://doi.org/10.37213/cjal.2021.28992

Rehner, K., & Mougeon, R. (2003). The Effect of Educational Input on the Development of Sociolinguistic Competence by French Immersion Students: The Case of Expressions of Consequence in Spoken French. *The Journal of Educational Thought / Revue de La Pensée Éducative*, *37*(3), 259-281. http://www.jstor.org/stable/23767258

EUROCALL 2023: CALL for All Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16995

Resourceful approaches in CALL for less-commonly taught languages (LCTLs): Case studies on Icelandic, Irish, and Nawat

Jane O'Toole[a], Branislav Bédi[b] and Monica Ward[c]
[a]School of Education, Trinity College, Dublin, otooleja@tcd.ie; [b]International Department, the Árni Magnússon Institute for Icelandic Studies, branislav.bedi@arnastofnun.is and [c]School of Computing, Dublin City University, monica.ward@dcu.ie

How to cite: O'Toole, J.; Bédi, B.; Ward, M. (2023). Resourceful approaches in CALL for less-commonly taught languages: Case studies on Icelandic, Irish, and Nawat. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16995

Abstract

This article explores challenges in supporting Less-Commonly Taught Languages (LCTLs) both in and outside of traditional classroom settings. The focus here is on three case studies of Icelandic as a second and foreign language (L2), Irish as a second or additional language (L2/ IAL), and Nawat as an endangered language. The linguistic contexts of the featured languages may differ- but all three have common ground in being less digitally resourced. Common practices in Computer-Assisted Language Learning (CALL) have shown that limited access to digital resources not only hampers development of CALL tools but also marginalizes pedagogical efforts in teaching and learning this language. In spite of this, CALL practitioners have shown resilience in adapting existing resources and developing new tools, while involving learners and speakers in collaborative efforts in CALL for LCTLs. Here, we showcase innovative approaches in piloting CALL applications for learning L2 Icelandic, integrating CALL into Irish language pedagogy and practice at English-medium primary level, developing game-based engines for Irish learners, and revisiting CALL resources for the endangered Nawat language. Our examples can offer valuable insights and inspiration to empower other CALL practitioners in supporting LCTL educators and learners, and indigenous and endangered language communities.

Keywords: *CALL, Less-Commonly Taught Languages (LCTLs), endangered languages, less resourced languages, collaborative efforts.*

1. Introduction

The term LCTLs refers to less frequently studied languages by students at North American universities[1], a definition which has been widely adopted to describe the teaching of languages other than French, German, and Spanish that are by this definition considered as More Commonly Taught Languages (MCTLs) (Gor & Katz, 2009). While its application may vary across countries and continents, we understand LCTLs as languages that are not frequently taught in a particular geographical region or a country (Ward, 2016). Both the spoken Irish and Icelandic languages enjoy official language status in their respective countries but are considered LCTLs in

[1] https://carla.umn.edu/lctl/definition.html

the wider learning context. For instance, the population of Iceland is relatively small (390,000 people) whereby English is widely spoken as an additional language of communication. The efforts to teach L2 Icelandic are hindered by relatively limited CALL resources in the local context of an increasing multinational population, who represent 16.3% of the total population. In Ireland, 40% of the population claim knowledge of the Irish language (1.8 million people) but only approximately one third of the population therein speaks Irish daily (Central Statistics Office, 2023). This includes *in* the education system, and the ever-decreasing figure of 71,968 of daily speakers of Irish *outside* of the education system (ibid.) is an ongoing cause for concern. The number of native speakers of the Nawat language in El Salvador has been estimated as less than 100 (Ward & Genabith, 2003), with more conservative estimates of 20 speakers proposed (ibid.). Nawat preservation and revitalization faces a range of complex challenges reflective of its critically endangered status.

Learners of these three languages face a mutual challenge of having limited access to digital resources. Some CALL resources may no longer be available or limit access to educators and learners (Gor & Vatz, 2009, p. 238), which may adversely affect resource visibility and the development of further digital resources. Our paper outlines resourceful approaches taken to support the teaching of Icelandic, Irish, and Nawat languages using various online tools and digital courseware. The specific examples of EL CALL projects and tools outlined in this article are informed by the collective research interests and collaborative review of the authors.

2. Facilitating and developing CALL resources for learners of Icelandic, Irish, and Nawat

2.1. Resourceful approaches in CALL for L2 Icelandic

To avoid digital minoritisation of Icelandic, i.e. when certain properties of the majority language disappear and characteristics of a minority language such as English emerge due to changes in society and technological advancements[2], the Icelandic government decided to strategically support language technology development (Nikulásdóttir et al., 2020) by launching a five-year programme in 2019. This programme aims at developing open access, open-source digital resources, and software to cater to the increased need for new technologies, such as Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) systems, machine translation tools, and grammar spell checks, to name a few. Similarly, other language technology projects and language tools, such as text and speech corpora, the database of Icelandic inflections DIM[3], to name a few, represent a crucial technological turning point in supporting the development of new and the enhancement of current CALL-related tools for both L1 and L2 Icelandic.

Accessibility to digital language resources and CALL tools for L2 Icelandic for children is still limited. For adult L2 learners, the most popular tool is *Icelandic Online*[4], which has reached over 80,000 users over the past 20 years (Arnbjörnsdóttir et al., 2020). This tool has undergone different technical and content improvements and is now offered in three different delivery modes: free, paid course with a tutor, and a hybrid course used in combination with in-classroom teaching. Moreover, this technology can be adapted to assist with L2 Icelandic literacy courses for children. To other CALL projects for teaching L2 Icelandic belong, the Learning And Reading Assistant – LARA[5] (currently being reimplemented with ChatGPT-4 as a software component) platform, the grammar-training application TunguPal[6], and the pronunciation training application CAPTinl[7]. All

[2] The MOLICODILACO Project: https://molicodilaco.hi.is/description/#deumert14

[3] https://bin.arnastofnun.is/DMII/

[4] https://icelandiconline.com

[5] https://lara-portal.unige.ch/view/index.php

[6] https://tungupal.is

[7] https://captini.tullius.dev/index

of these facilitate the learning of different language skills online. When exploring the Internet, 36 different online tools supporting the learning of L2 Icelandic have been found (Bédi, 2022).

Developing resources for CALL in L2 Icelandic poses various challenges. The main one is securing funding to guarantee steady work and completion of projects. Governmental and/or private funding bodies have partially or fully supported the projects and tools mentioned above. Researchers' and practitioners' own funding and time have also been used here. Enthusiasm at both individual and governmental level is thus a combined driving force behind supporting the development of digital resources for CALL in L2 Icelandic. The most recent example of a successful initiative taken by the Icelandic government is the established collaboration between OpenAI and private industry to use GPT-4 and create resources to promote the preservation of the language in the digital era[8]. Collaboration with local and international research, and educational consortia, has proved very important in sharing experience and knowledge among researchers, practitioners, and policy makers in CALL-related fields.

2.2. Resourceful approaches in CALL for Irish as an Additional Language (IAL)

This section introduces two Irish language CALL studies that contributed to the development of online/virtual learning domains for learners of Irish in a primary education setting, where limited opportunity to engage with Irish exists for learners.

A sociocultural theory-informed Irish language action research study introduced a Class Online Learning Zone (COLZ) as one of its piloted pedagogical approaches in order to engage senior Irish language learners. Utilising an established blog platform in primary education, the teacher-researcher curated a closed language learning area accessible by the teacher-researcher, children, and parents over the course of 10 weeks. The study outlines: (i) encouraging engagement levels overall; (ii) positive disposition of learners to the platform and to individualized and asynchronous aspects of the language exchange; (iii) a preference for goal-oriented and CLIL language activities; and (iv) active engagement of learners with teacher feedback online. In addition, learner partcipation with COLZ also supported engagement with a class Twitter (now X) account (O'Toole & Devitt, 2022)- reflective of how blog-based platforms can act as a gateway to engage with other platforms. A positive disposition towards goal-oriented computer-based activities was reflective of other school-based studies, including those specific to the Irish language (Dalton & Devitt, 2016). Finally, it was apparent that the COLZ platform provided an authentic domain for children to engage with the Irish language. It complemented the class Twitter account post writings by enabling students to publish their work for their teacher and parents only. Written Irish language exchanges in the form of feedback and informal posts constituted a new form of language interaction between teacher and child which extended beyond the classroom. Technology-related difficulties which arose for students included device access, parental engagement, and navigating Irish language orthography and diacritics for the first time.

Incorporating a sociocultural framework, a VR interactive game named *Cipher: Faoi Gheasa* [Cipher: Under a Spell] aims to foster student engagement and to provide enriched learning activities that empower student agency and choice (Xu et al., 2023). This initiative exemplifies how existing Natural Language Processing (NLP) resources can be adapted to introduce new educational tools for LCTL learners (ibid.). By repurposing the English version of 'Cipher' and integrating it with an established Irish language tool, 'Gramadóir' [Grammar Checker], the project effectively leverages available resources, offering cost-effective LCTL tools for educators and students. Preliminary feedback from senior primary students and teachers has been positive, underscoring the tool's potential integration into Irish education (ibid.). This proactive engagement with primary schools signifies the commitment of CALL developers to transition university-based NLP research into practical applications for Irish language education, maintenance, and revitalization.

Overall, both Irish language CALL studies contribute to the development of online/virtual learning domains for learners of Irish; which is particularly relevant to Irish language teaching and learning in English-medium education, where Learners' limited opportunity to engage with Irish is apparent. The Digital Plan for the Irish

[8] https://openai.com/customer-stories/government-of-iceland

Language (Government of Ireland, 2022) five-year programme can further contribute to and support the development of additional CALL resources for all Irish language learners and speakers (e.g. speech and text corpora to support pronunciation and grammar training, and spell checker systems). Its successful implementation will require partnership between commissioned speech and language technology project leaders, the education sector, and policy makers.

2.3. Resourceful approaches in CALL for the endangered language Nawat

The development of CALL courseware for speakers and learners of the Nawat language for the Pipil community in El Salvador (Ward, 2002; Ward & Genabith, 2003) is revisited. The additional challenges faced by the indigienous Pipil community speakers of Nawat are explored with reference to the tragic history of its speakers, economic challenges, critically endangered status of the language, and cultural sensitivities (ibid.). Within this complex cultural context, clarity, transparency, and ease of access to CALL resource development is essential both on a practical and cultural level. It is thus recommended that the development of CALL resources in this context is lean in terms of system requirements, low-cost (or ideally free), and can enable rapid-prototyping. Relatedly, it is also suggested that the developed courseware "be clear, consistent, reliable, easy to use, and be available in multiple modalities" (Ward & Genabith, 2003, p.248).

Drawing on Hubbard's (1996) Methodological Framework, the Nawat language courseware incorporated XML-based in order to develop web-based, CD-based, and printed language CALL materials in addition to a draft a syllabus which could be adapted as necessary (Ward & Van Genabith, 2003). The sustainable model supports an incremental development of additional lessons and courseware when and if required (ibid.). As we acknowledge the Decade of Indigenous and Endangered Languages 2022-2032 (UNESCO, 2021), this project highlights the critical importance of CALL resource development in relation to EL language documentation, community development, language publicity, and language prestige.

3. Conclusions

The three case studies presented here raise several important points for reflection. Across all three languages, the commonalities are as follow:.

- The importance of engagement of researchers and practitioners to develop CALL resources and to ensure open and free access to the target language learners
- The utilization of established programmes and courseware for MCTLs and leveraging this knowledge for the benefit of LCTL/EL learners

The differences in all three LCTLs are as follows:

- The lack of state support in the case of the Nawat language compared to the five-year programmes for developing digital language resources for Icelandic and Irish
- The development of digital resources specifically for children in Icelandic and Irish but not necessarily in Nawat
- The potential negative impact of globalization in the case of Nawat in contrast with globalization and migration as a possible driver of Icelandic and Irish language learner engagement

We can conclude that initiatives to collaborate not only with learners and the target language community groups, but also with local and international CALL research and educational teams, can stimualte useful insights and positively affect the development of resourceful CALL deployment across learning contexts.

References

Arnbjörnsdóttir, B., Friðriksdóttir, K., & Bédi, B. (2020). Icelandic online: twenty years of development, evaluation, and expansion of an LMOOC. In K.M. Frederiksen, S. Larsen, L. Bradley & S. Thouësny

(Eds), CALL for widening participation: short papers from EUROCALL 2020 (pp. 13-19). Research-publishing.net. https://doi.org/10.14705/ rpnet.2020.48.1158.

Bédi, B. (2022). Development of online tools supporting the learning of Icelandic as a foreign and second language. In B Bédi, H.J. Þorláksdóttir & K. Friðriksdóttir (Eds), *Perspectives on language and context, festschrift in honor of Birna Arnbjörnsdóttir* (pp. 47-56). Reykjavík: Stofnun Vigdísar Finnbogadóttur í erlendum tungumálum. https://vigdis.hi.is/wpcontent/uploads/2022/03/birnaarnbjornsdottir.afmaelisrit.mars2022.pdf.

Central Statistics Office. 2023. *Census of population 2022: Summary Results*. https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/educationandirishlanguage/

Coppens, X., Rico, M., & Agudo, J. E. (2013). Using blogs: authentic material and ranking quality for SLA. *The EUROCALL Review*, *21*(1), 20-34.

Dalton, G., & Devitt, A. (2016). Irish in a 3D world: Engaging primary school children. *Language Learning & Technology*, *20*(1), 21–33. doi: http://dx.doi.org/10125/44440.

Gor, K., & Vatz, K. (2009). Less commonly taught languages: Issues in learning and teaching', in M.H. Long & C.J. Doughty (Eds), *The handbook of language teaching* (pp. 234-240). Oxford, UK: Wiley-Blackwell.

Government of Ireland. (2022). *Digital plan for the Irish language: Speech and language technologies 2023-2027*. Dublin: The Stationery Office.

Nikulásdóttir, A., Guðnason, J., Ingason, A. K., Loftsson, H., Rögnvaldsson, E., Sigurðsson, E.F., & Steingrímsson, S. (2020). Language technology programme for Icelandic 2019-2023. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 3414–3422). Marseille, France: European Language Resources Association. https://aclanthology.org/2020.lrec-1.418

O'Toole, J., & Devitt, A. (2022). Twitter as a dynamic language learning platform for learners of Irish as an additional language (IAL) in a primary school setting in Ireland: Review and recommendations. In Arnbjörnsdóttir, B., Bédi, B., Bradley L., Friðriksdóttir, K., Garðarsdóttir, H., Thouësny, S., & Whelpton, M.J. (Eds), Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022 (pp. 302-307). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1475

UNESCO. (2021). Global action plan of the international decade of indigenous languages, abridged version.

Ward. M. (2016). CALL and less commonly taught languages – still a way to go. In S. PapadimaSophocleous, L. Bradley & S. Thouësny (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 468-473). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.608

Ward, M. (2002). Reusable XML technologies and the development of language learning materials. *ReCALL*,*14*(2), 283–292.

Ward, M., & Genabith, J. (2003) CALL for endangered languages: Challenges and rewards. *Computer Assisted Language Learning*, *16*(2-3), 233-258. http://doi.org/10.1076/call.16.2.233.15885

Xu, L., Ward, M. and Uí Dhonnchadha, E. (2023). Exploring the synergies between technology and socio-cultural approaches in CALL for less commonly taught languages. *10th Language & Technology Conference, Poznan, Poland, 21-23 April.*

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16992

Using Pimsleur for the self-regulated learning of spoken phrases in Brazilian Portuguese: a case study

Francis Will[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, francis-will@hotmail.com and [b]Department of Education, Concordia University, walcir.cardoso@concordia.ca

How to cite: Will, F.; Cardoso, W. (2023). Using Pimsleur for the Autonomous Learning of Spoken Phrases in Brazilian Portuguese: A Case Study. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16992

Abstract

This case study examines the self-regulated use of Pimsleur, a Language Learning Platform (LLP), as a tool to aid in the acquisition of spoken phrases in Brazilian Portuguese (BP) and their related pronunciation. Like many LLPs, research on Pimsleur is scant, as is the number of studies done on BP compared to other major languages. This study aims to address this gap in research. The participant-researcher completed the Pimsleur program through daily study over a 10-week period, after which quantitative data were collected through a post-test and delayed post-test. The results showed that Pimsleur contributed to the learning of the target phrases in the short term and that the participant produced speech that was highly intelligible, moderately comprehensible, but heavily accented. This shows that Pimsleur can be an effective tool for developing spoken BP and can offer a unique learning experience with its methodology and mobile capability that mitigates some of the issues around mobile-assisted language learning (e.g. app attrition).

Keywords: *Brazilian Portuguese, self-regulated learning, Pimsleur, language learning platforms, pronunciation, vocabulary.*

1. Introduction

In recent years, language learning beyond the classroom has been significantly affected by the rising popularity of technology-enhanced Language Learning Packages (LLPs). One such LLP is Pimsleur, which is at the center of this study. With rising accessibility, LLPs have multiplied in recent years, giving learners a considerable selection of apps to study, each marketing distinct curricula and expected learning outcomes for their users (Scacchi, 2022). The emergence of LLPs has had a positive impact on language learning (Loewen et al., 2019), particularly among learners interested in setting their own learning goals and monitoring and regulating their progress. For instance, a study by Loewen et al. (2020) involving students learning Spanish autonomously using Babbel (an LLP) showed that this app is effective for increasing explicit vocabulary knowledge (for similar patterns, see Vesselinov & Grego, 2019 and Rachels & Rockinson-Szapkiw, 2018). Despite these positive results, LLPs are problematic due to their limited interaction potential and the negative attitudes learners have toward them, which can lead to lower motivation and high attrition rates (e.g. Tuncay, 2020).

Pimsleur is based on an auditory teaching method developed by Pimsleur called *graduated interval recall*, a system whereby the app teaches the learner a linguistic item via translation and then recalls the item with decreasing frequency in subsequent lessons (Pimsleur, 1967). Pimsleur translates the words/phrases so that the

learner knows their meaning and uses native speakers so that the user knows how the word is pronounced. It promotes the long-term retention of linguistic items by using multiple, spaced exposures instead of a single study session, which is an effective practice strategy for SLA (Choe, 2016). According to Chapelle (2003), repetition is an input enhancement technique that is fundamental in SLA and, in the case of vocabulary, research shows that word knowledge and recall are optimized by meeting a word ten or more times (Webb, 2007). In their study on the efficacy of Pimsleur with Spanish learners, Vesselinov and Grego (2019) found that the app was effective in increasing oral proficiency and that user satisfaction was high after using the LLP.

Pimsleur was chosen for this research for two main reasons. Firstly, this study is interested in oral performance, and this LLP distinguishes itself from other popular apps because it concentrates on *spoken* phrases, with little to no consideration for the explicit teaching of grammar or literacy skills (Pimsleur, n.d.). Secondly, this study focuses on a lesser-taught language (BP) in a foreign learning context, where learners have limited access to the target language and its speakers. According to Nation (2001), an L2 learner in this situation will benefit considerably from having the kind of enriched input that Pimsleur offers.

As such, this study focuses on two underinvestigated topics: Pimsleur as a pedagogical tool and Brazilian Portuguese as the target language, a less commonly taught L2. This study aims to address this gap by investigating the effect of the self-regulated use of Pimsleur as a pedagogical tool in the learning of 36 spoken phrases in BP, including their pronunciation. This study addressed the following Research Questions (RQs):

RQ1: To what extent can a learner acquire 36 target phrases in Brazilian Portuguese after completing Pimsleur's curriculum in a self-regulated context?

RQ2: Based on holistic measures of pronunciation (intelligibility, comprehensibility, accentedness), to what extent can the learner be understood by speakers of BP after using Pimsleur in a self-regulated context?

2. Method

2.1 Participant and materials

The participant in this case study was the first author, a 33-year-old male with no previous knowledge of BP and no previous experience using Pimsleur. The rater, the second author, was a native speaker of BP. Pimsleur's BP program is comprised of 150 lessons across five levels with each level containing 30 lessons. Each lesson took the participant approximately 40 minutes to complete, with the participant completing the entire curriculum over a ten-week period. With roughly two hours per day dedicated to Pimsleur, the total time spent was approximately 100 hours of self-regulated BP study. For a description of Pimsleur's lesson scheme, see Frumkes (2021, p. 260-261)

Pimsleur lessons consist of 30-minute audio sections focused on rote repetition, followed by additional exercises designed to reinforce learning by linking spoken language with written language (which improves learners' grapheme-to-phoneme associations), and providing focus-on-form practice. The participant spent approximately 30 minutes on these activities, followed by 10 minutes dedicated to completing each lesson's practicing activities such as 'quick match' (a multiple-choice exercise), 'flash cards' (e-cards with a word on one side and its English translation on the other side), and 'speak easy' (listen and repeat).

2.2 Instruments

This study explores the extent to which a language learner can acquire 36 phrases in BP (see Appendix) and the extent to which these phrases can be aurally understood by a speaker of BP, based on three measures of pronunciation: intelligibility (the extent to which a listener actually understands an utterance), comprehensibility (the listener's belief of how difficult it is to understand an utterance), and accentedness (the extent to which a person's speech sounds compared to that of the target variety); for details on these measures, see Derwing and

Munro (2005). The phrases selected for this study were adapted from Baumgartner and Thering (2022), which are useful for general communication, socializing, working, and shopping. They are also covered in Pimsleur's curriculum.

Since the participant had no prior experience with or exposure to BP, no pretests were administered. A post-test was conducted immediately upon completion of the Pimsleur program, with a delayed post-test following two weeks thereafter. During this period between tests, the participant did not engage with BP through practice or exposure.

Vocabulary (RQ1): The instruments for measuring vocabulary and pronunciation were adapted from van Lieshout and Cardoso (2022). Accordingly, the computation of learning for the 36 target phrases was measured by the researcher, who rated the accuracy of each utterance using a 0-1 scale: 1 point was given for correct phrases, 0.5 for partially correct (e.g. incorrect use of a word or expression), and 0 for fully incorrect phrases.

Pronunciation (RQ2): The instruments used to measure intelligibility, comprehensibility, and accentedness were based on Derwing and Munro (2005). The rater was given a scalar judgment task to measure the comprehensibility and accentedness of the participant's speech using a 9-point Likert scale (1=completely incomprehensible/very accented; 9=completely comprehensible/not accented at all). Intelligibility was measured by having the rater listen to the recordings of the target phrases and transcribe what was heard. The researcher then evaluated the proportion of words transcribed correctly. For comparison purposes, the results for the transcribed phrases were converted into a 9-point scale.

2.3 Procedure

Upon completion of the Pimsleur program, the participant was audio-recorded producing the target 36 phrases in Portuguese immediately after the intervention (post-test) and two weeks later (delayed post-test). Pronunciation was evaluated by an L1 native speaker of BP who rated (for comprehensibility and accentedness) or transcribed (for intelligibility) all 36 audio samples.

3. Results

3.1. Vocabulary

To determine the extent to which a learner can acquire 36 target phrases in BP (RQ1), the rater assessed each utterance and determined whether it was correct, partially correct, or incorrect. The results in Table 1 show that the participant made significant gains in vocabulary retention, going from no knowledge to scoring 27/36 points (75%) on the post-test. However, there was a significant drop in the number of phrases retained on the delayed post-test, with the participant only scoring 24.5/36 (68%). These findings indicate that the initial learning gains achieved after completing the Pimsleur program are not fully sustained long-term.

Table 1. Vocabulary Gains: Mean Scores

Measures	Phrases learned in BP
Pretest	N/A
Post-test /36	27
Delayed post-test /36	24.5

3.2. Pronuciation

Table 2 illustrates the data on how understandable the participant's pronunciation was when speaking the target phrases (RQ2), considering the three measures adopted. On the post-test, the participant had a mean score of

8.7/9 (98%) in comprehensibility, 4/9 (44%) in accentedness, and 8.8/9 (99%) in intelligibility (note: higher scores indicate better performance). On the delayed post-test, the speaker had a mean score of 6.8/9 (76%) in comprehensibility, 3.5/9 (39%) in accentedness, and 8.8/9 (98%) in intelligibility. These results indicate that the participant had a considerable foreign accent when speaking, but remained intelligible (this measure remained stable) and relatively comprehensible (decreased by approximately two points) after completing the Pimsleur curriculum.

Table 2. Learner Pronunciation: Mean Scores

Measures	Comprehensibility	Accentedness	Intelligibility
Pretest	N/A	N/A	N/A
Post-test /9	8.7	4	8.8
Delayed post-test /9	6.8	3.5	8.8

4. Discussion

The considerable gains in phrases by the learner indicate that Pimsleur can be an effective learning tool for acquiring commonly used phrases. The results also suggest that shortly after using Pimsleur, it is possible for learners to produce speech that is highly intelligible, considerably comprehensible, yet quite accented. Finally, these results indicate that learners using Pimsleur in a self-regulated learning context can make important gains in spoken Brazilian Portuguese, consistent with previous findings about Pimsleur's ability to teach oral skills (e.g. Vesselinov & Grego, 2019). The results also support previous studies that have shown that a speaker can be highly intelligible despite having a heavy accent (Derwing et al., 2008).

The results of this study are also significant insofar as they partially respond to a limitation in van Lieshout and Cardoso's (2022) study using Google Translate, which examined this LLP's ability to assist in the self-regulated learning of Dutch phrases and pronunciation. In their study, there was no assessment of pronunciation during the delayed post-test because many participants forgot some of the vocabulary during the period between the post-test and the delayed post-test. To address this issue, the researchers called on future research to use a longitudinal model of self-regulated learning over a sustained period with systematic, spaced repetition. While our study does not focus on Google Translate, it does meet much of the criteria laid out in van Lieshout and Cardoso's call for future research. While the participant in our study also experienced a loss in vocabulary following the post-test, this was not a major issue as it did not impede him from producing phrases that could be rated for pronunciation.

5. Conclusions

The goal of this case study was to start addressing a gap in research involving the use of LLPs for the self-regulated learning of spoken Brazilian Portuguese. This study has shown that Pimsleur can effectively teach the autonomous learner useful phrases in BP that can be understood by native speakers in the short term.

It should be noted that this case study has some limitations. First, there was only one participant; a future study with a greater number of participants would give a more diverse and dependable sample of language learners. The fact that this study used only one rater is also an issue; a future study using additional raters would produce results with greater validity.

The Pimsleur mobile application coupled with the rote audiolingual method bypasses some of the challenges of many mobile-assisted language learning applications and enables substantial learning, as evidenced by the retention of the target 36 phrases (vocabulary) and related pronunciation in Brazilian Portuguese.

References

Baumgartner, M. & Thering, R. (2022, June 24) 40 Basic English phrases you'll use over and over. Fluent-U. https://www.fluentu.com/blog/english/common-english-phrases

Chapelle, C. (2003). *English Language and Technology*. John Benjamins. https://doi.org/10.1075/lllt.7

Choe, A. T. (2016). A critical review of Pimsleur language learning programs. *Hawaii Pacific University TESOL Working Paper Series*, *14*, 108-120.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, *39*(3), 379-397. https://doi.org/10.2307/3588486

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied linguistics*, *29*(3), 359-380. https://doi.org/10.1093/applin/amm041

Frumkes, L. A. (2021). Pimsleur. *CALICO Journal*, *38*(2), 259-270. https://doi.org/10.1558/cj.41955

van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning & Technology*, *26*(1), 1-19. http://hdl.handle.net/10125/73460

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, *31*(3), 293-311.

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, *53*(2), 209-233. https://doi.org/10.1111/flan.12454

Nation, I. S. P. (2001). *Learning vocabulary in another language* (Vol. 10). Cambridge: Cambridge university press. https://doi.org/10.1017/CBO9781139524759

Pimsleur (n.d.). The Pimsleur® method. Retrieved September 23, 2022, from https://www.pimsleur.com/the-pimsleur-method.com/the-pimsleur-method

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal*, *51*(2), 73-75. https://doi.org/10.1111/j.1540-4781.1967.tb06700.x

Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement andself-efficacy. *Computer Assisted Language Learning*, *31*(1-2), 72–89. https://doi.org/10.1080/09588221.2017.1382536

Scacchi, M. (2022, March). *Installs of Language Apps in Europe Surpassed 20 Million in Q1 2022.* Sensor Tower. https://sensortower.com/blog/state-of-language-apps-europe-2022

Tuncay, H. O. (2020). *App Attrition in Computer-Assisted Language Learning: Focus on Duolingo* (Doctoral dissertation, McGill University (Canada)).

Vesselinov, R., & Grego, J. (2019). Pimsleur efficacy study: Final report. Retrieved June 29, 2020, Pimsleur_EfficacyStudy2019.pdf (comparelanguageapps.com).

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, *28*(1), 46-65. https://doi.org/10.1093/applin/aml048.

Appendix

Language Objectives (36 spoken phrases)

1. Hi! I am [Name]. (And you?)	**1. Olá! Eu sou [Nome]. (E você?)**
2. Nice to meet you.	**2. Prazer em conhecê-lo.**
3. Where are you from?	**3. De onde você é?**
4. What do you do?	**4. O que você faz?**
5. What do you like to do (in your free time)?	**5. O que você gosta de fazer (no seu tempo livre)?**
6. What is your phone number?	**6. Qual é o seu número de telefone?**
7. Do you have coffee?	**7. Você tem café?**
8. Thanks so much.	**8. Muito obrigado.**
9. I really like it.	**9. Eu realmente gosto.**
10. Excuse me.	**10. Com licença.**
11. I am sorry.	**11. Sinto muito.**
12. What do you think?	**12. O que você acha?**
13. That sounds great.	**13. Isso parece ótimo.**
14. (Oh,) never mind.	**14. (Oh) não importa.**
15. I am learning Portuguese.	**15. Estou aprendendo português.**
16. I do not understand.	**16. Não entendo.**
17. Could you repeat that please?	**17. Você poderia repetir isso, por favor?**
18. Could you please talk slower?	**18. Você poderia falar mais devagar?**
19. Thank you. That helps a lot.	**19. Obrigado. Isso ajuda muito.**
20. What does 'do not enter' mean?	**20. O que significa 'não entre'?**
21. How do you spell that?	**21. Como você soletra isso?**
22. What do you mean?	**22. O que você quer dizer?**
23. Can you help me?	**23. Você pode me ajudar?**
24. I am looking for my wallet.	**24. Estou procurando minha carteira.**
25. Where is the bathroom?	**25. Onde fica o banheiro?**
26. I do not know.	**26. Não sei.**
27. I need a glass of water.	**27. Preciso de um copo de água.**
28. Where can I find this?	**28. Onde posso encontrar isso?**
29. How much does this/that cost?	**29. Quanto custa isso/aquilo?**
30. I do not need your help, thanks.	**30. Não preciso de sua ajuda, obrigado.**
31. Can someone help me do this?	**31. Alguém pode me ajudar a fazer isso?**
32. How can I help you?	**32. Como posso ajudá-lo?**
33. I will be with you in a moment.	**33. Estarei com você em um momento.**
34. What time is our meeting?	**34. Que horas é nossa reunião?**
35. Please call me (back) at 514 739 0826	**35. Por favor, me ligue (de volta) em 514 739 0826**
36. (Oh really?) Actually, I thought I was right.	**36. (Ah, é mesmo?) Na verdade, achei que estava certo.**

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16983

Widening the lens on technology-mediated Irish language learning policy and practice to enable Irish language, technology, and education policy alignment, integration, and enactment at school level

Jane O'Toole[a]

[a]School of Education, Trinity College, Dublin, otooleja@tcd.ie

How to cite: O'Toole, J. (2023). Widening the lens on technology-mediated Irish language learning policy and practice to enable Irish language, technology and education policy alignment, integration and enactment at school level. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16983

Abstract

Learner engagement with the Irish language in English-medium education (EME) in Ireland has been an area of concern in recent years (Inspectorate, 2022; Devitt et al., 2018). The lack of opportunity for learners to experience Irish as a living language outside of school exacerbates the situation. Technology-mediated language learning can provide learners with much-needed opportunities to consolidate Irish language learning as well as engage with the language across a range of authentic contexts and develop language learner networks. As the International Decade of Indigenous Languages 2022-2032 unfolds, this paper explores both Irish language education and digital education policy (Department of Education, 2022) at primary level within the broader context of national Irish language policy. Engagement with the Irish language at primary EME level is firstly explored in parallel with current digital learning policy at school level. Irish language (Government of Ireland, 2010; 2018) and Irish language digital policy (Ní Chasaide et al., 2022) at national level are then examined with reference to Irish language education. The exploration of Irish language policy domains seeks to address the question: how can the development of technology-mediated language learning be optimised in the context of Irish language teaching, learning, and engagement in the EME primary sector? Recommendations to support (technology-mediated) Irish language learning at EME primary level are suggested with reference to potential policy alignment and enactment.

Keywords: *Irish language education policy, digital learning policy, primary school, learner engagement.*

1. Introduction

The precarity of the Irish language is an ongoing concern in Irish society and the domain of education remains central to Irish language revitalisation efforts. An official language of the state along with English, the Irish language also bears the complex status of also being a native, heritage, and endangered language. The most recent census figures (Central Statistics Office, 2023) indicate a further decrease in native Irish speakers who use Irish outside the education system (71,968- a decrease of 1,835 since 2016) while reporting an increase in the number of people who claim some level of competency in the language (an increase of 6% between 2016 and 2022). Despite growth and development of and the Irish-immersion Education (IME), the Irish language continues to struggle in English-Medium Education (EME) which accounts for 92% of primary schools in the

Republic of Ireland (Gaeloideachas, 2023). Attainment in the Irish language amongst English-medium primary school goers has witnessed a steady decline since the 1980s and out-of-school use opportunities remain challenging to find and to facilitate. It is argued that technology-mediated language learning in endangered language contexts, such as Irish, can provide learners with dedicated opportunities to expand both their language learning and learning networks.

As Figure 1 illustrates, primary education policy is central to the point of departure of this policy review. Recent policy and practice is firstly examined with a focus on: (i) Irish language engagement; and (ii) the role of technology in primary education. National Irish language policy is then explored in order to identify common ground and alignment to further support the development of technology-mediated learning of Irish at primary level. The broader context of endangered languages and less commonly taught languages is then considered as a further lens for policy development. Finally, a series of recommendations are made in relation to education policy and practice in order to progress and enhance learners' opportunities to experience more integrated technology-mediated learning with respect to the Irish language at primary level in English-medium education.

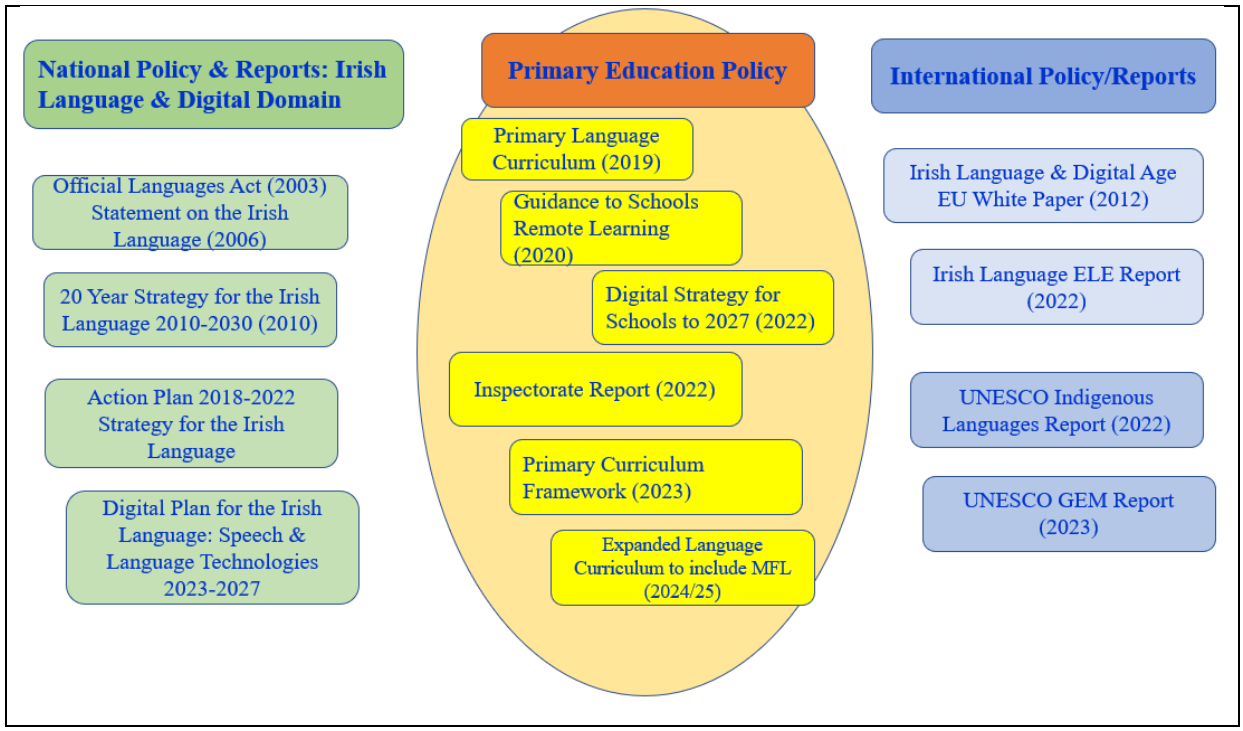


Figure 1. Outline of relevant education and language policy domains

2. Irish language learners and digital learning at primary level

Children's attainment in the Irish language in primary EME has been in decline in recent deacdes (Harris, 2008), coupled with excess disengagement with the language in comparision to other subjects (e.g. Maths, Literacy) (Devitt et al., 2018). The most recent Inspectorate[1] report (Inspectorate, 2022) identified the teaching and learning of Irish language at primary level as an overarching key concern clearing stating that "pupils' learning outcomes, motivation, and engagement in Irish need to be improved" (Inspectorate, 2022, p.105). It was noted that during the 2016-2020 period of inspections, students' attainment of learning objectives was good or very good in only 67% of lessons evaluated during notified inspections, and that the quality of learning outcomes during unannounced inspections was good or very good in 72% of instances; both figures highlight the potential to improve the quality of children's learning experiences. In addition, it was outlined that children's learning of

[1] The Inspectorate is a division of the Department of Education which coordinates and leads the evaluation of primary and secondary schools and education centres.

Irish had "suffered disproportionately" during the pandemic (Inspectorate, 2022, p.115) which further exacerbates Irish language efforts.

Notably, it was also identified that teachers should "make greater use of digital technology to support relevant and meaningful learning experiences in the teaching and learning of Irish" (Inspectorate, 2022, p.114). More generally, scope to enhance and embed children's engagement with digital technologies across the curriculum was identified, whereby only 54% of pupils surveyed reported use of digital technology in some lessons. While the role of technology in primary education in Ireland can be further enhanced, it is nonetheless readily acknowledged that schools' development of technological platforms (Department of Education, 2020) was accelerated and enhanced during the Covid-19 pandemic (Inspectorate, 2022).

3. Primary education digital learning policy

The Digital Strategy for Schools to 2027 (Department of Education, 2022) identifies digital learning as a key component and enabler for curriculum reform. Drawing on education stakeholder consultations, the need for "existing and new curriculum specifications to provide clear opportunities for digital teaching, learning, and the assessment giving teachers a clear rationale for using digital technology" (Department of Education, 2022, p.8) in primary and secondary education is acknowledged. The Digital Strategy places digital learning as an integrated and essential component of children's learning experience across the primary curriculum with a supporting focus on the importance of teacher professional learning to facilitate this. The importance of embedding technology in teaching and learning for the purpose of nurturing digital competence of students is clearly communicated. Direct links with the potential of digital platforms and engagement for Irish language learning, however, are not explicit. This is understandable given the broader remit and scope of the Strategy which, similar to the Primary Curriculum Framework (National Council for Assessment and Curriculum (NCAC), 2023), references subject groups such as literacy, language, the arts, and STEM, for example, as opposed to individual subjects. It is observed that preface of the Digital Strategy references how the Strategy itself will be instrumental in supporting developing curricular areas, such as Modern Foreign Languages (MFL) and STEM; thereby referencing the 'shininess' of the proposed forthcoming MFL element without reference to the huge potential technology-mediated learning holds for Irish language learning and languages of the school community.

The recently launched Primary Curriculum Framework (NCCA, 2023) aligns with the Digital Strategy and its predecessor by identifying *being a digital learner* as one of seven key competencies for development amongst school goers. As Figure 2 indicates, *being a communicator and using language* is also recognised as a competency across the curriculum.

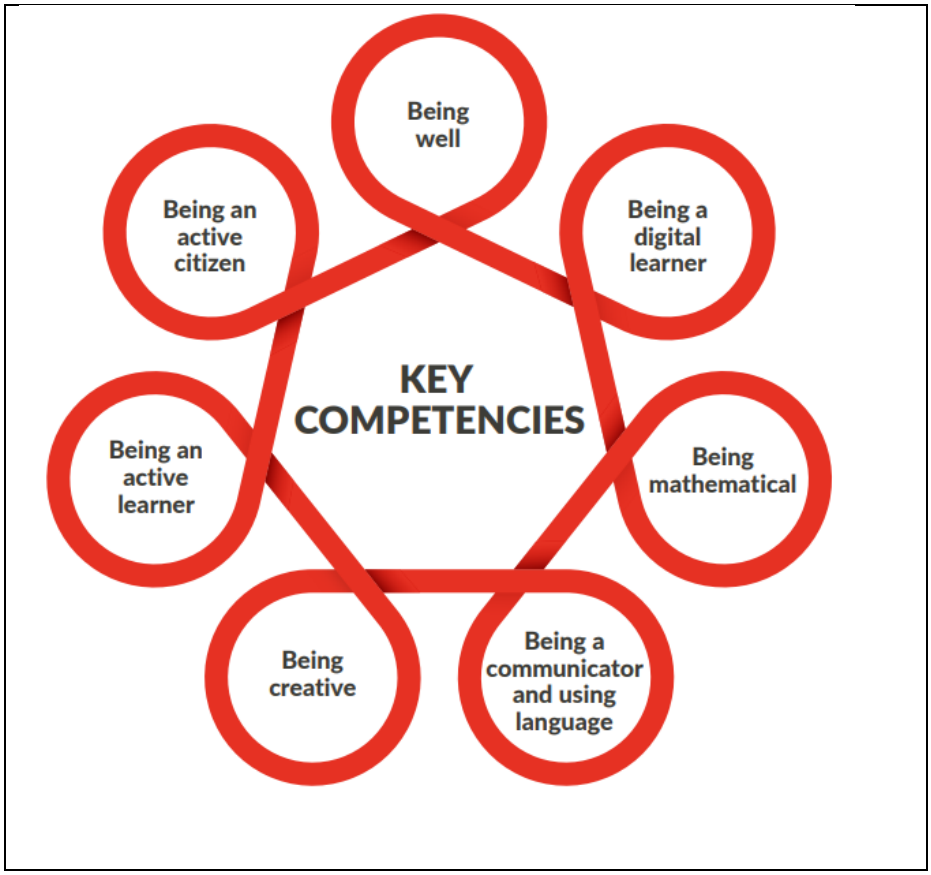


Figure 2: Key competencies for learners at primary level (NCCA, 2023, p.8)

The importance of digital literacy is reiterated in the Digital Strategy which in turn relates to the languages taught, learned, and spoken at school. Furthermore, alignment with the Primary Languages Curriculum (2019), which proposes an integrated approach to the teaching and learning of English and Irish, is outlined and acknowledges how digital literacy is set out therein as an important element of children's language learning in the two languages of the school (Department of Education, 2022).

4. Irish language and technology policy at national level

This section explores the broader realm of: (i) national Irish language policy; and (ii) Irish language technology policy, in order to situate Irish language and technology education policy in the national (and European) context. What efforts are being made at national level in both sectors which can potentially bolster technology-mediated Irish language learning at primary EME level?

At national policy level, the noughties witnessed significant progress for Irish language policy with the launch the Official Languages Act (2003), followed by the Statement of the Irish Language (2006) which paved the way for the 20 Year Strategy for the Irish Language (Government of Ireland, 2010).

The 20 Year Strategy for the Irish Language (2010) was initially a source of hope for learners of Irish given its detailed objectives and projections; however, the enactment of many aspects of the strategy has been widely considered ineffective. For example, the objective which seeks to increase the number of people with knowledge of Irish to two million is ill-defined and does not address contexts of use of the language nor competency.

The more recent Action Plan for the Irish Language 2018-2022, which details the implementation of the Strategy's objectives, arguably indicates the relative lack of integration between the domains of education and technology in the Irish language context.

Area for Action 1:	Education
Area for Action 2:	The Gaeltacht
Area for Action 3:	Family Transmission of the Language – Early Intervention
Area for Action 4:	Administration, Services and Community
Area for Action 5:	Media and Technology
Area for Action 6:	Dictionaries
Area for Action 7:	Legislation and Status
Area for Action 8:	Economic Life
Area for Action 9:	Cross-Cutting Initiatives

Figure 3: Areas for Action in Action Plan for the Irish Language 2018-2022 (Government of Ireland, 2018)

As the structure of the Action Plan indicates, Education and Media and Technology are discreet entities. Closer examination of the actions for Education not only indicate a low proportion of actions specific to Irish language learning in English-medium education (e.g. development of a programme to support the Primary Languages Curriculum (PLC), pilot programme of partial Irish language immersion utilising CLIL), but also an absence of proposals in relation to technology-mediated Irish language learning. Reference to the provision of digital resources by Foras na Gaeilge is made explicitly for Irish-medium schools. Media and Technology actions are detailed with many spanning various domains including television, radio, the arts, digital, technology, and education. The latter focussed on third level provision. The final Area for Action 9 Cross Cutting Initiatives does not include integration of the domains of education and technology, which arguably represents an opportunity missed to align and meaningfully integrate both domains in order to further support learners of Irish.

Preceding the Language Strategy's Action Plan (2018), a White Paper (Judge et al., 2012) made recommendations for further investment and resourcing by the state to support Irish language Speech and Language (S&L) technologies, in addition to strongly advocating for the potential role of technology for Irish language learners. Ten years later, these views reflected in the European Language Equality (ELE) Report (Lynn, 2022), which also detailed the deficient status of Irish language digital infrastructure and technologies compared to other EU languages. The urgent need to develop a range of core technologies was outlined in order for the Irish language to keep pace digitally in an increasingly globalised world. While primarily addressing the fundamental advances required in the fundamental S&L technologies, the potential impact on the education sector was also highlighted citing the acute dearth of CALL tools for the language, which in turn results in the provision of outmoded Irish learning activities for students (Lynn, 2022).

The ELE report (Lynn, 2022) set the context for the long-awaited Digital Plan for the Irish Language (Ní Chasaide et al., 2022). The plan references the Adapt Centre at Dublin City University and the Phonetics and Speech Laboratory at Trinity College, Dublin, as the principal partners of commissioned research including speech synthesis technology and beta speech recognition technology system development. Significantly for education, Chapter 14 of the plan provides an extremely comprehensive account of the current speech and language technology both under development and currently utilised in the broader education domain, in addition to providing a strong rationale for the role of Irish language technologies in education in Ireland. It is of note, however, that the commissioning and implementation of the Digital Plan for the Irish Language resides with the Department of Tourism, Culture, Arts, Gaeltacht, Sport, and Media which has a separate remit to the Department of Education.

5. Conclusion: Potential steps forward in policy and practice to further facilitate technology-mediated Irish language learning in primary EME

As the Digital Plan proposes, "the symbiosis of powerful language technologies and digitally available knowledge of Irish structure – along with how it is acquired – have the potential to bring about a paradigm shift in Irish language education" (Ní Chasaide et al., 2022, p.18). However, the exploration of the policy areas of Irish language education, digital learning, Irish language, and Irish language technology development suggest a

lack of connection and integration. In order to facilitate a graduated integration of Irish S&L technologies at (primary) school level, it is recommended to further inter-departmental cooperation and collaboration at government level (between the Department of Education, and the Department of Tourism, Culture, Arts, Gaeltacht, Sport, and Media, for example). The establishment of a dedicated working group inclusive of all relevant education, Irish language, and digital learning stakeholders could support such collaboration. The acknowledgement, endorsement, and circulation of Chapter 14 of Digital Plan for the Irish Language by the Department of Education to all schools could simply begin a conversation with practitioners about the potential of Irish language S&L technologies in the classroom. The Inspectorate could support schools in showcasing effective practice in terms of Irish language teaching and learning drawing on S&L technologies as observed in evaluations to date. The involvement of professional support services for teachers and school leaders[2] and expert partners in integrating S&L technologies into dedicated Continuing Professional Development (CPD) and support for schools, informed by the commissioned Irish language research emanating from the Digital Plan in the broader context of endangered language education (UNESCO, 2021) and technology in education (UNESCO, 2023).

References

Central Statistics Office (CSO). (2023). *Census of population 2022: Summary results*. Retrieved at: https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/educationandirishlanguage/

Department of Education. (2020) *Guidance on remote learning in a COVID-19 context: September – December 2020, For primary schools and special schools*. Retrieved at: https://assets.gov.ie/90252/070614f4-3dff-4489-8e4d-be8cd625999e.pdf

Department of Education. (2022). *Digital strategy for schools to 2027*. Stationery Office. Retrieved at: https://www.gov.ie/en/publication/69fb88-digital-strategy-for-schools/

Devitt, A., Condon, J., Dalton, G., O'Connell, J., & Ní Dhuinn, M. (2018). An maith leat an Ghaeilge? An analysis of variation in primary pupil attitudes to Irish in the growing up in Ireland study. *International Journal of Bilingual Education and Bilingualism* 21(1) 105-117. doi: 10.1080/13670050.2016.1142498

Gaeloideachas. (2023, July 31). *Statistics* Gaeloideachas.Retrieved at: https://gaeloideachas.ie/i-am-a-researcher/statistics/#:~:text=Figures%20from%20the%20Department%20of,schools%20in%20the%2026%20counties

Government of Ireland. (2010). *20-year strategy for Irish language 2010–2030*. Dublin: Government of Ireland. Retrieved at: http://www.ahg.gov.ie.elib.tcd.ie/ie/Straiteis20BliaindonGhaeilge2010-2030/Foilseachain/Strait%C3%A9is%2020%20Bliain%20-%20Leagan%20Gaeilge.pdf

Government of Ireland. (2018). *Action plan for the Irish language 2018-2022: 20-year strategy for the Irish language, 2010-2030*, Retrieved at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/88777/193c4c89-d711-4498-b3b1-42cdbaedb0cd.pdf#page=null

Harris, J. 2008. The declining role of primary schools in the revitalisation of Irish. *AILA Review* 21(1) 49-68. https://doi.org/10.1075/aila.21.05har

[2] Oide (Irish language term for 'teacher') is the newly amalgamated support service for teachers and school leaders in Ireland funded by the Department of Education (www.oide.ie)

Inspectorate. (2022). *Chief inspector's report [2016-20]*. Dublin, Ireland: The Stationery Office. Retrieved at: https://www.gov.ie/en/publication/611873-chief-inspector-reports/

National Council for Curriculum & Assessment. (2023). *Primary curriculum framework*. Dublin: NCCA. Retrieved at: https://curriculumonline.ie/getmedia/84747851-0581-431b-b4d7-dc6ee850883e/2023-Primary-Framework-ENG-screen.pdf

Ní Chasaide, A., Lynn, T., Uí Dhonnchadha, E., Ní Chiaráin, N and Judge, J. (2022). *Digital Plan for the Irish Language: Speech and Language Technologies 2023-2027*. Dublin, Ireland: Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media. Retrieved at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/250129/1425436f-e1da-4661-8483-92d9ddb4a716.pdf#page=null

UNESCO. (2021). Global action plan of the international decade of indigenous languages, (IDIL2022-2032); abridged version. Retrieved at: https://unesdoc.unesco.org/ark:/48223/pf0000379853?posInSet=1&queryId=7ea9f278-dbc4-45fc-8a43-a3cceefecd3f

UNESCO. (2023). Technology in education: A tool on whose terms? Global education monitoring report. Paris: UNESCO. Retrieved at: https://unesdoc.unesco.org/ark:/48223/pf0000385723

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16873

Appropriating Zoom to host and conduct writing center tutorials

Chris Harwood[a]
[a]Faculty of Liberal Arts, Sophia University, Japan, , chrisharwood@sophia.ac.jp

How to cite: Harwood, C. (2023). Appropriating zoom to host and conduct writing center tutorials. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16873

Abstract

This paper explores the technological affordances of using Zoom for pedagogical purposes in the context of one-to-one writing center tutorials. First, the theoretical pedagogical framework that informs writing center tutorials is explained and the salient mediating affordances of Zoom technology are outlined. Then, tutor and student qualitative feedback from the beta testing of a virtual writing center hosted in Zoom are presented and briefly discussed. The feedback suggests that Zoom has a range of technological affordances that can be employed by educators to host and conduct meaningful writing center tutorials in an undergraduate EMI context.

Keywords: *videoconferencing, pedagogy, mediation, L2 writing, affordances.*

1. Introduction

In the mid-1990s, the advent of the Internet enabled university writing centers to offer asynchronous online services such as email-based tutoring and document exchanges. By the mid-2000s, as the technological revolution advanced, writing centers began to appropriate collaborative file editing applications such as Google Docs to offer synchronous text-based tutorials. In recent years, increases in bandwidth technology and the development of reliable videoconferencing applications such as Zoom have enabled writing centers to offer virtual tutorials with video and document sharing capabilities (Harwood & Koyama, 2021). Considering how the tools and features of Zoom videoconferencing technology can be appropriated for pedagogic purposes is important for engaging English as an additional language students in the writing process and supporting their learning in Virtual Writing Centers (VWCs). Therefore, this paper investigates the technological affordances of Zoom that can be employed by educators to host and conduct VWC tutorials.

Prior to analyzing the technology under focus, it is useful to briefly explain the theoretical framework that informs many writing center tutorial practices. A cornerstone of writing center pedagogy is sociocultural theory (Nordlof, 2014), which asserts that learning is social in nature. According to sociocultural theory, cognitive development is dependent on the interpersonal interactions a person engages in and the cultural tools they appropriate as they learn.

A person's independent capabilities can be enhanced by providing incremental support (or scaffolding) to help them achieve their learning goals. The support should gradually decrease as the learner's competence increases, until they can complete the learning objective(s) independently (Vygotsky, 1978). As Wells (1999) contends, dialogic learning of this kind promotes an interactive learning process, interweaving the individual and their societal context. These theories underpin the dialogic teaching style in writing centers, characterized by ongoing, reciprocal exchange between tutors and students that is intended to foster active engagement and critical thinking.

2. The Pedagogical Affordances of Zoom

A Zoom meeting room is a virtual space where videoconferencing takes place. It is the virtual equivalent of a physical conference room, where participants can join to engage in a group discussion or presentation via video and audio. Zoom has a feature which allows the host to divide the main meeting room into smaller groups named *breakout rooms* for more focused discussions or group work (see Figure 1).

The host (or VWC administrator) can then bring meeting participants back into the main meeting room when needed. Zoom also has a *waiting room* feature which automatically places participants in a separate virtual space when they join the meeting. This enables VWC administrators to control tutorial entry to the VWC and provides a level security as administrators can monitor who enters the meeting. The feature includes a sound notification, akin to a doorbell, alerting VWC administrators when a participant is waiting to join the meeting.

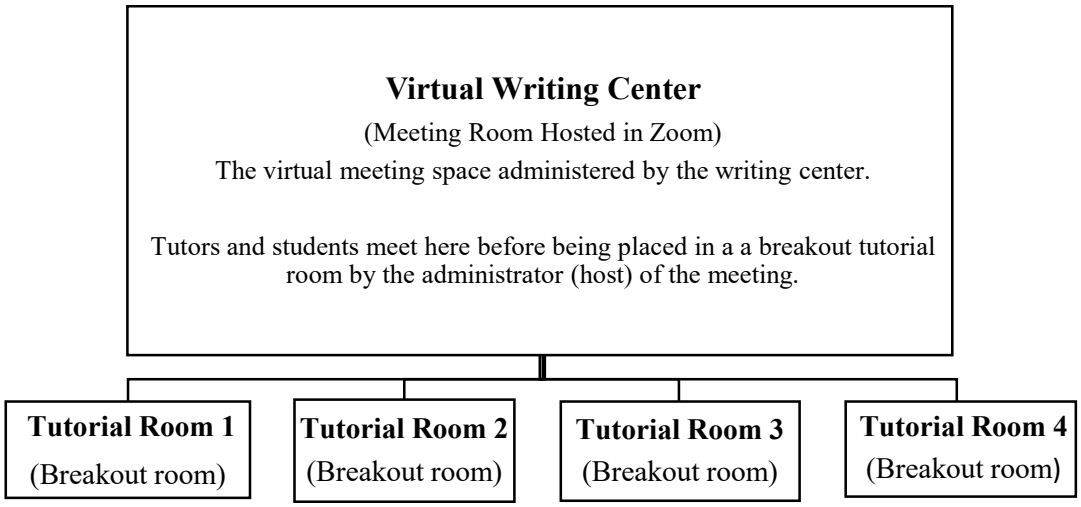


Figure 1. A visual representation of a Zoom VWC

Zoom offers various other audio-visual features that are potentially useful for pedagogical purposes and writing center administration. The key tools and features of Zoom that can be appropriated for pedagogical purposes are described below.

- *Screensharing* facilitates real-time collaboration and feedback while preserving students' digital ownership of their work as they do not need to send a copy to the tutor.
- *Virtual whiteboards* serve as a creative space for brainstorming, breaking down complex concepts, organizing ideas, and facilitating group problem-solving.
- *The chat function* enables tutors to record points of feedback and advice for future reference, as well as share links to additional resources for students. It is also a useful feature for language students, neurodivergent students, and students that are hard of hearing as they do not have to rely on audio clarity.
- *Automated captioning* can enhance understanding for language students and students that are hard of hearing by providing text subtitles.
- *The pinning feature* can highlight visual cues from the speaker, reducing distractions and aiding comprehension.
- *Virtual backgrounds* provide privacy for participants, minimize background distractions, and can promote professionalism e.g. VWC administrators and tutors can have the same writing center branded background.

According to Greer and Harris (2018), online writing instruction should adopt a user-centered design model that emphasizes student engagement. Indeed, much of the research on Zoom for teaching considers the pedagogical utility of its key features. Studies have found that students perceive breakout rooms as a beneficial learning space as they enable increased social interaction and teacher engagement, and that video, screen sharing, and annotations can promote interactive synchronous learning (Berges et al., 2021).

Furthermore, features such as chat, emojis, and whiteboards are reported as effective for increasing student engagement (Kohnke & Moorhouse, 2022). Studies have also reported that teachers find screen sharing to be especially useful for direct instruction and real-time resource sharing (Stevenson et al., 2022), and that Zoom can facilitate the learning of writing in EFL courses (Huong, 2023).

Despite these reported benefits, Zoom presents technological limitations and issues related to camera use. For instance, the lack of social context cues in Zoom can increase cognitive workload and lead to Zoom fatigue (Bailenson, 2021). Additionally, the constant use of video cameras can cause anxiety and fatigue among students (Ngien & Hogan, 2022). Thus, educators should be aware of these issues and adjust their practices accordingly.

3. Method

The current study is based on tutor and student feedback from the beta testing of a VWC introduced at Sophia University's Faculty of Liberal Arts (FLA) Writing Center in Tokyo, Japan. The center provides English writing support primarily for Japanese students with high English proficiency. Due to the COVID-19 pandemic, in 2020 the center moved its services online and used Zoom as a VWC.

3.1 Participants and procedure

17 FLA undergraduate students (English proficiency TOEFL 105+), who were members of the same university club, tested the VWC by connecting to it via Zoom and engaging with tutors in breakout rooms. The students then collectively evaluated their experiences and documented them in a report. The students also completed post-tutorial reflection forms in which they individually responded to the question: *Reflect on your tutorial and the advice you were given. How will you integrate your thoughts and the tutor feedback in your writing?* Five writing center tutors also provided feedback through a 45-minute focus group conducted on Zoom. The tutors were asked to reflect on their experiences of using the VWC and share how they used the technology. The tutor and student names in the feedback below are pseudonyms.

The VWC was considered beneficial and efficient for tutorials by both tutors and students. It is also noteworthy that the VWC staff successfully scheduled, coordinated, and administered the Zoom breakout room writing center tutorials. They reported that the waiting room feature helped them manage the flow of students and tutors entering the VWC, especially during the changeover times between tutorial sessions. No substantial issues related to the affordances of the Zoom technology were reported by the VWC staff.

3.2 Tutors' feedback on the VWC

Tutors reported that they appropriated several features of Zoom breakout rooms to enhance their interactions with students and provide feedback. All five tutors noted how useful the screen sharing function was for focussing on the student's writing and for facilitating the understanding of paragraph organization and coherence, paraphrasing, and in-text citations.

Three of the five tutors mentioned that they had to explain to several students about modifying the permissions settings in Zoom. This was because tutors need students to give them permission to 'control the screen' and scroll through the students' writing so they can provide feedback through track-changes. The whiteboard feature was used by three tutors to visually explain ideas and concepts linked to students' writing and research. The chat feature was employed by all five tutors to share notes about student writing and hyperlinks to writing resources. Chat was also used to troubleshoot technical issues when audio and camera issues occurred as students worked out how to navigate the platform. Although tutors had to initially assist students navigate the VWC, they perceived Zoom as very useful for conducting writing tutorials, noting that it was not a cause of struggle during sessions. As Shunsuke, a writing center tutor in the focus group, explained:

> "The Zoom tutorials were fairly similar to face-to-face tutorials in terms of outcomes…I had good sessions face-to-face and had some bad session face-to-face, and I had good sessions online and problematic sessions online. I don't think that was related to the technology but more the student themselves… whether they are passive or disengaged and so on…"

Although the VWC offers a tutorial experience different from face-to-face tutorials for tutors, the technological affordances of Zoom clearly provide a virtual space that facilitates purposeful tutorial interactions with students.

3.3 Student feedback on the VWC

Student feedback from the post-tutorial reflection forms indicates that the students were actively engaged in improving their writing skills in the VWC. Table 1 below shows the number of student participants that made references to various aspects of their writing during the beta testing of the VWC.

Table 1. Aspects of the writing process students mentioned they engaged in the VWC tutorials

Aspect of writing	Number of students
Recording notes	12
Clarifying vocabulary	10
Recognizing objectivity	7
Revising to enhance clarity	5
Revising to enhance organization	8
Citation	11
Paragraph structure	10
Coherence	10
Categorization	4
Logical thinking	6

Students noted that they used chat to record notes and clarify the meaning of vocabulary. They also noted that they had begun to recognize the significance of objectivity in academic writing and the necessity of revising to enhance clarity and organization. Furthermore, students reported that they addressed a variety of topics in the tutorials, including citations, paragraph structure, logical thinking, coherence, and categorization. A salient observation regarding the student's post-tutorial reflections is that no technology issues were mentioned. Moreover, all students provided details of writing issues discussed during the VWC tutorials and mentioned how they were planning to revise their writing (Harwood, 2023). Three examples of student post-tutorial reflections are given below in Table 2 below.

Table 2. Examples of students post-tutorial reflection comments

Name	L1	Post-tutorial reflection
Hana	Japanese	The tutor advised me to include more outside sources to make my statements more reliable and accurate. I will also change some of the sources I used since it is a little bit outdated for the topic I am writing about. I will keep in mind to use more recent sources. My essay was neither a block pattern nor a point-by-point pattern so I will change that and put the counterargument and rebuttal only in the third paragraph. As I am going to change my last paragraph, I will also have to change part of my thesis statement.
Yuki	Japanese	I was able to go over the overall structure and content of the essay that I initially wanted to cover. I also went over some referencing issues. I will integrate the feedback that I got into my writing by focusing on the introduction and conclusion sentences.
Cao	Chinese	I had problems with my writing regarding structure, relevance, and clarity were solved. The advice I got from the tutor regarding approaches to my topics were helpful. I will be making changes to my draft based on the feedback I got.

The student feedback illustrates that the VWC tutors were able to successfully leverage ZOOM functionalities to deliver engaging and interactive tutorial sessions to provide precise and actionable feedback.

4. Summary and Conclusion

This paper has provided a rationale for how the technological affordances of Zoom can be appropriated by university writing centers to offer one-to-one writing tutorials. Preliminary findings from the beta testing of a VWC along with tutor and student feedback show how Zoom can function as an instrumental tool in facilitating university writing center tutorials. The technology allowed tutors and students to engage in purposeful dialogues about the writing process in the VWC. Moreover, the student reflections of VWC tutorials suggest that Zoom provides a space to reflect on and evaluate their work, which also aids in recognizing aspects of their writing that require revision. As VWCs continue to be opened by university writing centers, research is necessary to evaluate how videoconferencing applications such as Zoom, and its evolving technological features, can be appropriated for pedagogical purposes. Accordingly, as technological advancements persist in shaping VWCs, tutors should regularly review and revise their VWC tutorial practices to ensure they fully exploit the technology for pedagogical purposes.

References

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of ZOOM fatigue. *Technology, Mind, and Behavior*, *2*(1). https://doi.org/10.1037/tmb0000030

Berges, S., Martino, S., Basko, L., & McCabe, C. (2021). "ZOOMing" into engagement: Increasing engagement in the online classroom. *Journal of Instructional Research*, *10*, 5–11. https://eric.ed.gov/?id=EJ1314160

Greer, M., & Harris, H. S. (2018). User-centered design as a foundation for effective online writing instruction. *Computers and Composition*, *49*, 14–24. https://doi.org/10.1016/j.compcom.2018.05.006

Harwood, C. (2023). Appropriating Zoom to Provide Access to One-to-One Writing Support. In J. Keengwe (Ed.), *Handbook of Research on Innovative Frameworks and Inclusive Models for Online Learning* (pp. 194-214). IGI Global. https://doi.org/10.4018/978-1-6684-9072-3.ch010

Harwood, C., & Koyama, D. (2020). Creating a virtual writing center to support self-regulated learning. *Studies in Self-Access Learning Journal*, *11*(3), 164–186. https://doi.org/10.37237/110306

Huong, L. P. H. (2023). Mediated learning of the writing skill via ZOOM by EFL students. In Bansal, R., Singh, R., Singh, A., Chaudhary, K., & Rasul, T. (Eds.), *Redefining virtual teaching learning pedagogy* (pp. 359–369). Wiley & Sons.

Kohnke, L., & Moorhouse, B. L. (2022). Facilitating synchronous online language learning through ZOOM. *RELC Journal*, *53*(1), 296–301. https://doi.org/10.1177/0033688220937235

Ngien, A., & Hogan, B. (2022). The relationship between ZOOM use with the camera on and ZOOM fatigue: Considering self-monitoring and social interaction anxiety. *Information, Communication & Society*, 1–19. https://doi.org/10.1080/1369118X.2022.2065214

Nordlof, J. (2014). Vygotsky, scaffolding, and the role of theory in writing center work. *The Writing Center Journal*, 45–64. https://www.jstor.org/stable/43444147

Stevenson, M., Lai, J. W. M., & Bower, M. (2022). Investigating the pedagogies of screen-sharing in contemporary learning environments: A mixed methods analysis. *Journal of Computer Assisted Learning*, *38*(3), 770–783. https://doi.org/10.1111/jcal.12647

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wells, C. G. (1999). *Dialogic inquiry*. Cambridge University Press.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16979

Beyond the walls of classrooms: Exploring the pedagogical effectiveness of Text-To-Speech-based Shadowing (TTS-S) on the development of Mandarin tones

Sue-Anne Richer[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, Montréal, Canada, , sue-anne.richer@mail.concordia.ca and [b]Department of Education, Concordia University, , walcir.cardoso@concordia.ca

How to cite: Richer, S.A.; Walcir, C. (2023). Beyond the walls of classrooms: Exploring the pedagogical effectiveness of text-to-speech-based shadowing on the development of Mandarin tones. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16979

Abstract

This study examines the pedagogical effectiveness of using Text-To-Speech synthesis (TTS) combined with Shadowing (TTS-S) for self-regulated learning of Mandarin tones 1 and 4. The aim is to determine the probability of success of this innovative approach that uses TTS to generate audio for shadowing practice. The research was guided by the following research question: can TTS-S help L2 learners raise their awareness and improve their perception and production of the target Mandarin tones over a period of six weeks? Over six week, ten participants engaged in self-regulated activities using TTS-S to learn the two pronunciation targets. By means of pre-/post-tests (to assess effectiveness in pronunciation) participants were asked to complete: (1) an awareness task in which they verbalized their metacognitive knowledge of Mandarin tones; (2) ABX tasks to assess their perception of Mandarin tones; and (3) a production task to evaluate their production of the target tones. Our findings are inconclusive regarding the effectiveness of TTS-S for improving awareness, perception, and production of Mandarin tones 1 and 4 among L2 learners. They also indicate that while certain aspects of phonological development, specifically production, showed some improvements, the overall impact was not statistically significant.

Keywords: *Text-To-Speech synthesis (TTS), shadowing, L2 pronunciation, speech perception.*

1. Introduction

With limited classroom time (Collins & Muñoz, 2016), teachers struggle to provide personalized language input (e.g. listening activities) and opportunities for students to practice oral output (speaking). Text-To-Speech synthesizers (TTS), also known as text readers, offer a possible solution by allowing students to interact with the computer anytime-anywhere, and at their own pace (Cardoso, 2022). As such, this technology has the potential to improve students' aural skills and provide flexible language practice (Little, 1995). Although TTS offers many benefits (e.g. immediate access to the language; Liakin et al., 2017), it lacks an output-inducing component that promotes active speech production (Fang, 2017). To address this issue and contribute to the field of computer-assisted pronunciation instruction, this study combines TTS with shadowing (i.e. the repetition of a word or phrases immediately after hearing it; Lambert, 1994), a technique that has been proven to be effective in developing L2 pronunciation (e.g. Foote & McDonough, 2017; Zajdler, 2020). By combining these two technologies, to which we will refer as 'TTS-based shadowing training' (TTS-S henceforth), our approach provides learners with the benefits of both TTS (exposure to input) and shadowing (opportunities to practice output).

To determine the probability of success of this innovative approach, this study examined the pedagogical effectiveness of using TTS-S in a self-regulated learning environment (whereby learners set goals for their learning and then attempt to monitor and regulate their progross) to acquire tones #1 and #4 in Mandarin Chinese. While tone #1 (high tone) is relatively easy to acquire in comparison with other tones, tone #4 (descending tone) is considered one of the hardest to produce (Hendry, 2023). The research was guided by the following research question: can TTS-S help L2 learners raise their sound awareness and improve their perception and production of the target Mandarin tones over six weeks? By means of pre-/post-tests (to assess effectiveness in pronunciation), ten beginner-level participants were asked to complete: (1) an awareness test to evaluate the participants' metacognitive knowledge of Mandarin tones; (2) two ABX tests to assess their perception of Mandarin tones; and (3) a production test to evaluate their ability to produce the target tones.

2. Method

2.1. Participants and procedures

Ten adult beginner learners of Mandarin participated in this study. The participants had no knowledge of the Mandarin tones and had never taken any Chinese courses. Prior to engaging in the study, participants were asked to sign an informed consent form and fill out a demographic questionnaire. Participants then completed the awareness test, the two ABX tests, and the production task. The participants were then provided with a brief description and analysis of the tones in Mandarin. For instance, they were told that tones are important in Mandarin because they contribute to meaning (e.g. they serve to differentiate words), and that they would discover what two of these tones are. The participants were then provided with five texts containing five to nine sentences in which the two targeted tones appeared a total of 66 times (tone #1 and #4 appear between 10 to 16 times a week). On a weekly basis (Weeks 1-5), the participants learned the target Mandarin tones by: (1) copying and pasting these texts into Ispeech (ispeech.org; the TTS application adopted); (2) listening to the synthesized output; and (3) immediately repeating it using the proposed shadowing technique (TTS-S). If they decided to use additional material, they were asked to add it to their weekly learning logs. As per the study's recommendation, it was expected that they completed one text every week. A weekly reminder was sent to the participants to indirectly track if participants complete the activities as instructed. The posttests were administered one week after the completion of the self-directed treatment proposed.

2.2. Instruments and data analysis

To evaluate the phonological development of tones, this study looked at the effects of the treatment on the sound awareness, aural perception, and oral production of tones in a self-directed TTS-S based environment. A sound awareness test was used to measure the metacognitive knowledge of the students about tones in Mandarin. Six minimal pairs (e.g. ma1 vs ma4) were played and the participants were asked to label the two sounds as the same or different and then, if applicable, explain how the sounds differ.

The aural perception tests used two ABX aural discrimination tests (for the rationale, see Hautus & Meng, 2002). In the first ABX tests, participants were asked to determine whether there was an audible difference between two audio signals. In the second ABX test, participants followed a match-to-sample discrimination procedure: two sounds (test and control) were presentend to the participants, followed by a third sound (blind sample) that corresponded to the test sound. The participants were asked to match the blind sample to the test item.

Because of the low proficiency of the participants and the short duration of the treatment, the production test adopted a 'listen and repeat' activity (for the rationale, see Cardoso et al, 2021), a less demanding test in comparison with reading aloud and spontaneous conversation. This test asked the participants to listen to a recording of a native speaker pronouncing a word in Mandarin and to repeat it as accurately as possible.

The phonological awareness data were analyzed via a 3-level system: 0 (fully incorrect = not aware); 1 (partially

correct = partial knowledge); and 2 (fully correct = full knowledge). The perception and production data were analyzed using descriptive statistics (e.g. means-M, standard devision-SD) via SPSS. All scores were converted to a 100% scale for comparison purposes. Means of pretests and posttests were compared using paired sample t-tests.

3. Results

Participants' awareness, perception, and production of tones were tallied to determine if the utilization of TTS-S within a self-directed learning context promotes the learning of the target Mandarin tones in terms of sound awareness, aural perception, and oral production. The paired sample t-test results indicate that the posttest means for all tests were higher compared to the means obtained in pretests. However, these differences were not statistically significant. Table 1 provides a summary of the descriptive results for the four tests administered: sound awareness (Awareness), ABX1 and ABX2 (Perception), and oral production (Production).

Table 1. Summary of the results.

Tests	Test (N=10)	Mean (%)	SD	Average Gains	SD
Awareness	Pretest Posttest	88.34 98.34	10.43 2.64	5	10.54
ABX1	Pretest Posttest	97 99	4.83 3.16	2	4.21
ABX2	Pretest Posttest	88 94.5	16.36 8.31	6.5	11.56
Production	Pretest Posttest	79.75 82	15.47 15.47	2.25	5.83

An interesting pattern observed across all tests is that some participants demonstrated improved knowledge of Mandarin tones (e.g. Participants 2, 6, and 10 in Figure 1), while others showed no change. Due to space limitation, we only provide the results for oral production, shown in Figure 1.

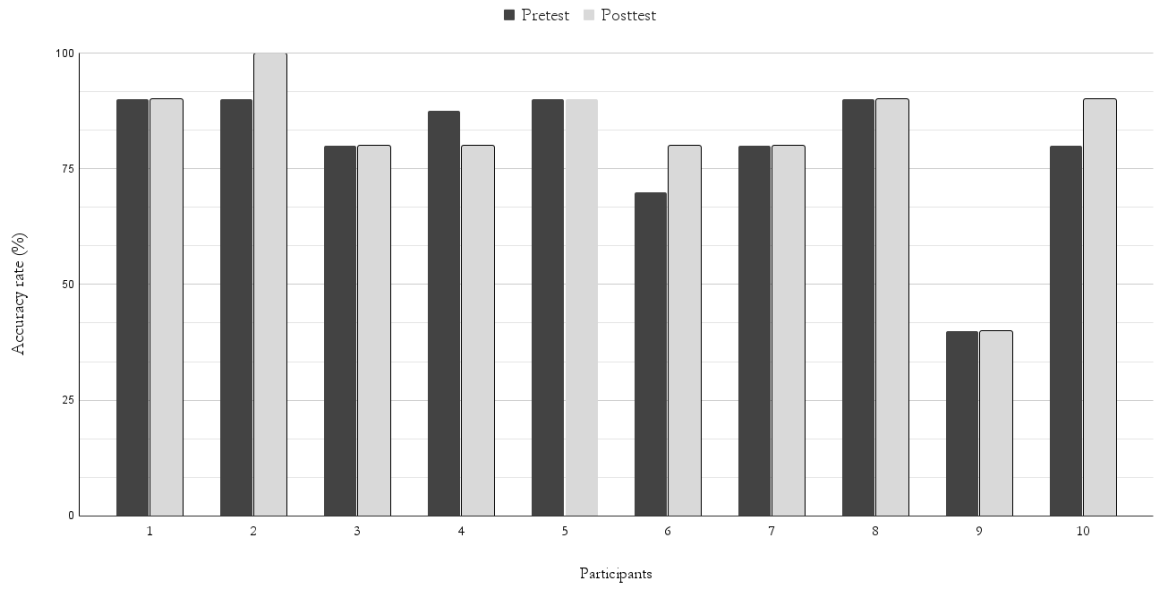


Figure 1. Production Task: Accuracy rate (%) by participant.

4. Discussion

To investigate the pedagogical effectiveness of a novel approach for self-regulated learning that combines TTS with shadowing for the learning of tones #1 and #4 in Mandarin Chinese, this study was guided by the following research question: can TTS-S help L2 learners raise their sound awareness and improve their perception and production of the target Mandarin tones over a period of six weeks? Results showed that the improvements observed in all four tests (i.e. sound awareness, aural perception, and oral production) were not statistically significant. This suggests that, overall, the proposed treatment did not yield any significant improvements in the assessed measures of pronunciation development.

Two important generalizations can be derived from our study. Firstly, our findings support the established notion that L2 phonology follows a trajectory that begins with the development of phonological awareness and aural discrimination (perception) and progresses towards the acquisition of oral production skills (Celce-Murcia et al., 2010; Schmidt, 2001). For instance, in general we observed that participants exhibited a high level of awareness and perception regarding tone 1 and tone 4, suggesting a near- or complete development. However, there remained room for improvement in the production of these tones, highlighting the sequential nature of tonal development where perception typically precedes production. Secondly, our study provides evidence that although certain aspects of tone development can be achieved within a semi-directed learning environment, as the one explored in this study via a combined use of TTS and Shadowing, the effectiveness of such interventions may vary among individuals due to inherent individual differences (Robinson, 2010). The varying degrees of improvement among participants highlights how personal factors can influence the effectiveness of our proposed intervention, with some benefiting more than others.

5. Conclusions

Based on the findings of this study, it is inconclusive to assert that TTS-S can effectively raise sound awareness or significantly improve the perception and production of the target two Mandarin tones among L2 learners within a six-week timeframe. The results indicate that while certain aspects of phonological development, specifically production, showed some improvements, the overall impact was not statistically significant. Importantly, the small sample size and the individual differences among the participants should be taken into consideration when interpreting these findings. Factors such as motivation and prior linguistic background may have influenced the outcomes. It is possible that some participants were more responsive to the TTS-S approach and demonstrated improvement, while others may not have benefited as significantly.

It is worth acknowledging some of the limitations of this study, which could also have influenced the results. A key limitation is its small sample size of only ten participants. Although this was done to ensure an in-depth examination of the participants' pedagogical experience (including learner perceptions, use of strategies in TTS-S/human interactions – not reported in this paper), larger samples and individual interviees are needed to obtain more accurate estimates and enhance the generalizability of the findings. A methodological limitation includes the simplicity of most tests (e.g. the first awareness test, ABX1, the listen-and-repeat oral production test), which were found to be extremely easy, with participants achieving ceiling or near-ceiling scores. Another limitation regards the lack of an account for individual differences among the participants, such as age, motivation, and learning strategies, which have been shown to significantly impact language learning (DeKeyser, 2013; Robinson, 2010). Further investigations are necessary to validate the pedagogical effectiveness of TTS-S and its suitability in language learning. In addition to the aforementioned directions for future research, it is important for replication studies to consider including all four tones in order to create a more authentic representation of tone use in Mandarin. A longitudinal approach would also provide more robust insights into the developmental path of Mandarin tone acquisition, and enable a better understanding of the relationship between language use and the various components that affect language development.

Acknowledgements

We would like to thank the ten participants for their time and dedication and the invaluable feedback provided by Dr. Kevin Papin. This research has been partially funded by a research grant from the Social Sciences and Humanities Research Council of Canada (430-2022-00512).

References

Cardoso, W., Collins, L., & Cardoso, W. (2021) Developmental Sequences in Second Language Phonology: Effects of Instruction on the Acquisition of Foreign sC Onsets. *Frontiers in Communication*, *6*, 662934. https://doi.org/10.3389/fcomm.2021.662934

Cardoso, W. (2022). Technology for speaking development. In T.M. Derwing., M.J. Murray., & R.I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 299–313). Routledge. https://doi.org/10.4324/9781003022497

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation*. New York: Cambridge University Press.

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal*, *100*(1), 133–147. https://doi.org/10.1111/modl.12305

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*(s1), 52-67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

Fang, C. (2017). *A Prototype Text Analyzer for Mandarin Chinese TTS sSystem* (Publication No. 1111826). [Masters' thesis, Uppsala University]. DiVa Portal. https://www.diva-portal.org/smash/get/diva2:1111826/FULLTEXT01

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1), 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Hautus, M. J., & Meng, X. (2002). Decision strategies in the ABX (matching-to-sample) Psychophysical Task. *Perception & Psychophysics*, *64*(1), 89–106. https://doi.org/10.3758/bf03194559

Hendry, C. (2023). *Using Translation Tools for L2-Learning in a Self-Regulated Environment* (Publication No 991740). [Doctoral dissertation, Concordia University]. https://spectrum.library.concordia.ca/id/eprint/991740/

Huang, Y., & Lawless, H. (1998). Sensitivity of the ABX discrimination test. *Journal of Sensory* Studies, *13*(2), 229–239. https://doi.org/10.1111/j.1745-459x.1998.tb00085.x

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

Little, D. (1995). Learning as dialogue: The dependence of learner autonomy on teacher autonomy. *System*, *23*(2), 175–181. https://doi.org/10.1016/0346-251X(95)00006-6

Robinson, P. (2010). Individual Differences and Instructed Language Learning. *Language Learning*, *60*(s2), 46-84. https://doi.org/10.1075/lllt.2

Schmidt, R. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction*, (pp.3-32). Cambridge University Press.

Zajdler, E. (2020). Speech shadowing as a teaching technique in the CFL classroom. *Lingua Posnaniensis*, *62*(1), 77–88. https://doi.org/10.2478/linpo-2020-0005

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.17001

Cognitive grammar hands-on: a design-based approach to the didactic integration of interactive grammar animations

Daniel Pust[a]

[a]Department of Germanic Studies, Otto von Guericke University of Magdeburg, daniel.pust@ovgu.de

How to cite: Pust, D. (2023). Cognitive grammar hands-on: a design-based approach to the didactic integration of interactive grammar animations. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.17001

Abstract

Tutorial computer-assisted language learning tends to employ a deductive teaching approach, explicating grammatical concepts prior to practice exercises. This leads to a didactic gap, impacting learner engagement. Interactive Grammar Animations (InGA) aim to bridge this gap by enabling learners to explore the conceptual motivation and meaningfulness of grammar on their own. To harness the potential of InGAs, the selection of a suitable learning object and its didactic integration are pivotal. Only against the backdrop of media-adequate didactics, InGAs grant alternative access to the learning content and can become a means of insight and understanding. To bridge both the didactic and the knowledge gap, the research project on InGAs investigates the functionality of the interactive learning medium and its integration into a didactic concept. As a case study, the project uses grammar animations on the German passive voice whose visual interface is extended so that language learners can manipulate the content presented within them. Following a Design-Based Research (DBR) approach, the project is currently in its fourth cycle and this article reports on the challenges of developing and integrating an interactive learning application that adopts an inductive approach to grammar instruction in the foreign language classroom and reflects on didactic as well as methodological aspects.

Keywords: *cognitive grammar, interactivity, task-based language teaching, didactic integration, design-based research.*

1. Extending the didactic repertoire of tutorial CALL

InGAs are an innovative learning medium with the potential to shift paradigms in terms of how Computer-Assisted Language Learning (CALL) is used within the classroom and beyond. Usually, tutorial CALL follows a deductive teaching approach to grammar by presenting explicit explanations of grammatical concepts before providing exercises for practice (Heift & Vyatkina, 2017). This didactic procedure puts a curb on the involvement and activity of the learner, as they are initially relegated to absorbing metalinguistic information. Moreover, the focus on formal features in subsequent exercises gives them little opportunity to practice skills related to semantic and pragmatic aspects of grammar despite being equally important for the development of grammatical competence (Purpura, 2004).

InGAs, on the other hand, overcome this didactic confinement by using an inductive approach and enabling learners to explore the conceptual motivation and meaningfulness of grammar (Langacker, 2008) on their own. Representing a further development from conventional grammar animations (Roche & Scheller, 2008), InGAs inherit the principal feature of using visual metaphors to explain abstract concepts of Cognitive Grammar to language learners in a way that is illustrative and easy to understand. But unlike their predecessors, the visual

display of InGAs is extended by an input interface, enabling language learners to interact with the elements displayed in the animations. As the learner actively shapes the animated content, a dynamic process of mutual influence between the learner and the digital learning medium provides a sophisticated form of interactivity that stimulates cognitive interactions between existing knowledge on the part of the learner and new impulses on the part of the application (Domagk et al., 2010).

However, to harness the learning potential of InGAs, the decisive factor is the selection of a suitable learning object and its didactic integration (Beatty, 2010). Only against the backdrop of media-adequate didactics, InGAs grant alternative access to the learning content and extend the didactic repertoire of tutorial CALL. To address this puzzle, the research project on InGAs adopts Task-Based Language Teaching (TBLT) as a didactic framework, and employs both quantitative as well as qualitative approaches of DBR (McKenney & Reeves, 2014) to monitor the design and development of the didactic concept. Currently, the project is in its fourth iteration of (re-)designing, testing, and evaluating within a case study on the German passive voice (cf. Roche & Suñer-Muñoz, 2016). Preliminary findings of this evolutionary process shed light on the challenges of developing and integrating an interactive learning application that adopts an inductive approach to grammar instruction in the foreign language classroom from a didactic as well as a methodological perspective.

2. Design-based research

In terms of research methodology, the study on InGAs is situated within the paradigm of DBR, an empirical research approach that aims to test and optimise the design of learning concepts, environments, and materials within a cyclical process. Rodríguez (2017, p. 365) summarises the features of DBR as follows, whilst the characteristics may vary from study to study: "synergy between design and research processes, iteration, methodological pluralism, intervention, exploration, and collaboration." Although this research approach seems to be particularly suitable for CALL research because of the strong link between theory and practice, the weaknesses of DBR, such as the lack of uniform methodological guidelines, the involvement of researchers, and problems with the generalisation of results, must also be taken into account.

To explore those factors in the complexion of the teaching-learning process that condition and influence the didactic functionality of InGAs, various interaction-centred methods are used to achieve a multi-perspective documentation with particularly rich data: user tracking to collect data on user behaviour; semi-structured interviews to obtain insights into the perceptions, feelings, and assessments of learners as well as of teachers; questionnaires to gather further information about the respondents and their language learning biographies; and work materials (participants' worksheets, the teacher's lesson plan as well as their reflection report).

To expedite the development of the didactic concept, three testbeds, differing in the degree of the ecology of the learning environment, were used (see Fig. 1). Testbeds differed as to the size and consistency of the learning group, the teacher, and observers sitting in on the language class. At the lowest level, the field of enquiry consisted of a particularly small learning group that was formed sporadically and instructed by the researcher in the role of the teacher; a more natural setting was when the researcher taught in an established language class with a regular practitioner as an observer. In future cycles, the researcher and the practitioner will switch positions. While the test environments varied, the characteristics of the test subjects were kept constant throughout. All of them were undergraduate or graduate students at Otto von Guericke University of Magdeburg with different cultural backgrounds, and consequently with different first languages, and language skills in German varying between A2 and B1 (CEFR).

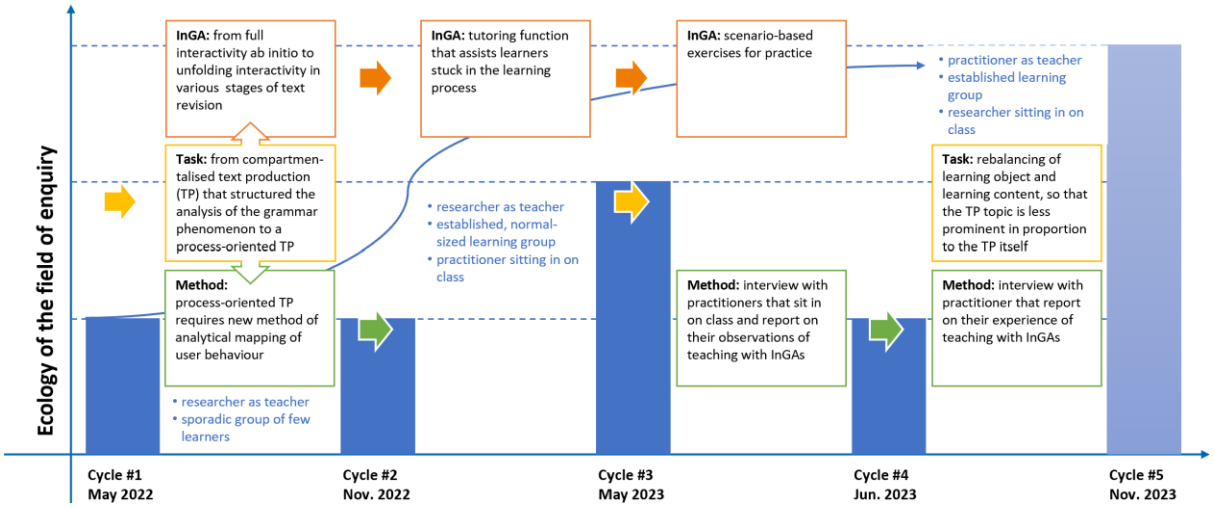


Figure 1. Evolution of the didactic concept for InGAs within four cycles of (re-)designing, testing, and evaluating (cycle #5 planned for the future at the moment of writing)

3. The evolution of a didactic concept for InGAs

3.1. TBLT

Given the exploratory approach to grammar afforded by the high degree of interactivity of InGAs, TBLT provides a didactic design that is ideally suited to engaging learners in active learning and meaningful language use mediated by technology (González-Lloret & Ortega, 2014). Tasks in language learning contexts can be understood as a framework within which learners carry out activities designed to achieve a goal through linguistic means (Van den Branden, 2006). While the design principles of TBLT have served as a foundation throughout the evolutionary development of the didactic concept for integrating InGAs, the tasks that guide the learning process underwent significant changes with the most notable modification after the first cycle.

For providing a linguistic activity that is authentic and oriented towards a communicative goal, learners were set the task of writing a protocol for a harmless animal experiment depicted in the animations (see Fig. 2). Initially, the text production was divided into three tasks each dealing with two of the six scenes of the animal experiment and each focusing on a different aspect of the diathesis in German. This compartmentalisation turned out adverse to the learning progress. Learners struggled with an analytical approach that focused primarily on meaning, and on form only in a second step. For this reason, the redesign of the didactic concept involved restructuring the text production from a product-oriented approach to a process-oriented approach, in which learners were guided through three stages of drafting and revising their texts. This new procedure has proved successful and the only adjustment that remains for the forthcoming fifth cycle is the balancing between the learning object and the learning content. This adjustment has turned out necessary as participants still find the prominence of the dilemma the animal faces within the experiment, which is intended to arouse their interest in the subject matter and prepare them for the subsequent virtual experiment, distracting from the grammatical phenomenon.

Figure 2. InGA screenshot of scene 1 of the animal experiment (version of cycle #4), conceptualised with the focus on the containers while the researcher and the pigeon are faded out, producing the verbalisation displayed. The bulbs next to the scene selector (left) represent feedback; the help button (right) provides hints on what to do next.

3.2. InGAs

To enable learners to cope with the challenging task of writing an experiment protocol, InGAs support them in composing the text. Since the support has to be adapted to the task demands, restructuring the writing task also meant adapting the functionality of the InGAs. This adaptation provided an opportunity for tackling another problem that occurred in the first cycle. While being helpful, the complex functionality of the various interactive elements can also be strenuous. Having all functions available to them as of the first task, participants felt initially overwhelmed in the first cycle, and the confusion caused by the complexity of the operations of the various parameters only dissipated as they worked through the tasks. But adopting a process-oriented approach for the text production in various stages made it also possible to introduce the functionality of the InGAs step by step, so that its complexity is gradually revealed to the learners.

While the interactivity of the elements in the animations enables learners to explore the relationship between the conceptual composition of a scene and its verbal realisation, thus providing them with suggestions for the wording of the experiment protocol, the InGAs are also equipped with a simple feedback system. This is meant to give learners orientation in the process of discovering different conceptualisations of the same scene and to alert them of ungrammatical sentences or sentences that are inadequate with regard to the communicative goal. However, as user behaviour continued to show difficulties in following learning paths after the second cycle, an additional support system was implemented to intervene when learners encountered obstacles and to provide hints on how to proceed. This intervention resulted in positive outcomes, as participants produced less unproductive input. Moreover, scenario-based exercises were introduced as of the fourth cycle to also give learners opportunities to practice the grammatical phenomenon by ‘helping’ avatars to complete their experiment protocol.

4. Adaptations of research methods

As a result of adjusting the tasks that facilitate learning with InGAs and the learning application itself in the development of the didactic concept, it was also necessary to adapt research methods accordingly. User tracking

is generally used to collect data on the learners' behaviour, providing information on the use of learning paths or deviations from them. Since learning paths depend on the structure of the task, restructuring the writing task after the first cycle also involved an adaptation of the analytical approach. Rather than looking at the participants' behaviour in performing a task on a particular scene, the process of drafting and revising the experiment protocol required a more extensive analysis of how they made use of InGAs to develop the text procedurally, compared to the text products.

A complementary source of data are interviews, which not only serve to gather information about the participants' perceptions, but also to obtain the expertise of practitioners for advancing the integration of InGAs, either as observers or as teachers depending on their role. Therefore, the interview guides for data collection have been constantly adjusted relative to the didactic arrangement and the different testbeds. Similarly, when interpreting the interview data by means of qualitative content analysis (Mayring, 2021), the categories for coding had to be repeatedly re-evaluated and, if necessary, adjusted. In circumstances of continuous optimisation, the flexibility of DBR allows for adapting research methods according to the variability of the object of investigation as well as of the field of enquiry.

5. Conclusions

Though the suitability of TBLT as a didactic framework for the use of CALL is by now an accepted fact and technology can also promote TBLT in turn (Lai & Li, 2011), the integration of InGAs in the foreign language classroom is more than just another positive example of the synergistic effects between these two fields of research. On the one hand, the inductive approach to grammar instruction with (semi-)open tasks shifts the paradigm of conventional tutorial CALL by expanding the didactic repertoire associated with it; on the other hand, it shows how digital learning media can be used to track each step of the learner's processing of the task. This in turn helps to achieve a better understanding of task in process in relation to task as workplan, as well as to align task demand and task support. Finally, the study provides evidence for the integration of a new CALL application using DBR through various stages of testing, evaluation, and redesign, and for making the best possible use of different testbeds. In the future, the InGA research project will advance by further testing, and potentially optimising, the didactic concept in various learning environments and with different types of learners. Another goal is also to test the effectiveness of InGAs in comparison with their non-interactive predecessors.

References

Beatty, K. (2010). *Teaching and Researching: Computer-Assisted Language Learning*. (2nd ed.). London: Routledge. https://doi.org/10.4324/9781315833774

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, 26(5), 1024-1033. https://doi.org/10.1016/j.chb.2010.03.003

González-Lloret, M., & Ortega, L. (2014). *Technology-mediated TBLT. Researching Technology and Tasks*. Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.6

Heift, T., & Vyatkina, N. (2017). Technologies for Teaching and Learning L2 Grammar. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (p. 26-44). Hoboken (NJ): Wiley Blackwell. https://doi.org/10.1002/9781118914069.ch3

Lai, C., & Li, G. (2011). Technology and Task-Based Language Teaching: A Critical Review. *CALICO Journal*, 28(2), 498-521. https://doi.org/10.11139/cj.28.2.498-521

Langacker, R. W. (2008). Cognitive Grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (p. 66-88). New York (NY): Routledge.

Mayring, P. (2021). *Qualitative Content Analysis. A Step-by-Step Guide*. London: SAGE.

McKenney, S., & Reeves, T. C. (2014). Methods of evaluation and reflection in design research. In D. Euler & P. F. E. Sloane (Eds.), *Design-Based Research* (p. 141-156). Stuttgart: Steiner.

Purpura, J. (2004). *Assessing Grammar*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511733086

Roche, J., & Scheller, J. (2008). Grammar Animations and Cognition. In F. Zhang & B. Barber (Eds.), *Handbook of Research on Computer-Enhanced Language Acquisition and Learning* (p. 205-219). Hershey (PS): IGI Global.

Roche, J., & Suñer, F. (2016). Metaphors and grammar teaching. *Yearbook of the German Cognitive Linguistics Association*, 4(1), 89-112. https://doi.org/10.1515/gcla-2016-0008

Rodríguez, J. C. (2017). Design-based Research. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (p. 364-377). Hoboken (NJ): Wiley Blackwell. https://doi.org/10.1002/9781118914069.ch24

Van den Branden, K. (2006). Introduction: Task-based language teaching in a nutshell. In K. Van den Branden (Ed.), *Task-Based Language Education: From Theory to Practice* (p. 1-16). Cambridge: CUP.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16980

Effects of online task-based peer interaction on learners' speech development and attitudes toward English as a lingua franca

Atsushi Iino[a] and Brian Wistner[b]
[a]Hosei University, Japan, iino@hosei.ac.jp and [b]Hosei University, Japan, wistner@hosei.ac.jp

How to cite: Iino, A.; Wistner, B. (2023). Effects of online task-based peer interaction on learners' speech development and attitudes toward English as a lingua franca. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16980

Abstract

This paper reports the outcomes of incorporating two kinds of videoconferencing into a university English as a Foreign Language (EFL) special seminar course in Japan for two semesters. One type of videoconferencing was a weekly interview with a non-native English speaker/instructor living outside Japan using a commercial online conversation program; the other was a weekly online peer meeting on Zoom where groups of three learners practiced speaking English using role play tasks. The effects of the two types of videoconferencing were examined through pre- and post-speaking tests and a questionnaire on the participants' attitudes toward English as an International Language (EIL; Nakamura, Lee, & Lee, 2018). Additionally, a qualitative analysis of peer-group interaction was conducted to reveal what happened in the peer-group role play task. The results indicated that speaking skills improved over time, and tolerant attitudes toward variations of English were observed. Abundant opportunities for automatizing EFL use and negotiation for meaning were observed in the peer-group role play.

Keywords: *videoconferencing, task-based instruction, EFL.*

1. Introduction

In EFL situations like Japan, learners of English have limited opportunities to use English orally inside and outside of educational institutions. The use of commercial online conversation services can provide opportunities for learners to improve their oral communicative competence and develop tolerant attitudes toward the variation of English (Iino, 2022). Another way is online peer interaction (Hetrovicz, 2021), which is claimed to be effective to promote intercultural communicative competence as well as second language (L2) oral proficiency (Warner-Ault, 2020).

The various aspects and benefits of in-person peer interaction were reviewed in Philp, Adams, and Iwashita (2013), and Adams and Oliver (2019). Regarding online peer interaction, few previous studies have focused on online interaction between Non-Native Speakers and other Non-Native Speakers (NNS-NNS), and many of them examined psychological changes. For instance, Lenkaitis (2020) examined how videoconferencing affected L2 Spanish learners' perceived autonomy over time and also included self-reported measures of L2 speaking skills satisfaction and L2 conversation length (i.e. the amount of time participants' felt they could converse in the L2). The quantitative results of the study indicated that L2 learners reported higher levels of autonomy, speaking

skills satisfaction, and L2 conversation length after participating in weekly L2 videoconferencing sessions with their classmates. However, gaps remain in the literature regarding the relative effectiveness of online interaction for L2 learning and the influence of interlocutors' linguistic and cultural background variables.

The purpose of the present study was to examine the effects of NNS-NNS interaction through videoconferencing with learners from different L1 and cultural backgrounds. Previous studies often relied on self-reported measures to assess L2 development and did not fully investigate the effects of exposure to international English on learners' attitudes toward varieties of English. Data based on standardized L2 speaking tests could provide insight into how learners' speaking skills quantitatively change over time when videoconferencing with a diverse group of English speakers. Thus, the following research questions were investigated:

1. To what extent does group-based videoconferencing in an English-as-a-lingua-franca setting improve L2 speaking skills?
2. To what extent does videoconferencing influence L2 learners' attitudes toward English as a lingua franca/international language?
3. What kind of negotiation for meaning occurs in task-based online peer-group role plays during videoconferencing?

2. Method

2.1. Participants

Nineteen university students majoring in economics at a university in Tokyo, Japan participated in this study. For two semesters, they took a special seminar to improve their oral English proficiency and intercultural communicative competence through an active learning cycle run by one of the researchers (see Figure 1). The group consisted of eight third-year students who had experienced one year of the learning cycle described below, and 11 second-year students who were new to the learning cycle. Their English ability was between CEFR A2 to B2, with the majority at the B1 level.

2.2. Treatment

The participants were directed to complete a role play task in groups on Zoom for about 30 minutes outside the class each week (online peer interaction; OPI). The task was practiced face-to-face in class beforehand in three-person groups, with each participant taking the role of a facilitator/decision maker, pro-side advocator, or con-side advocator. The learners were asked to repeat the same task with different class members online and to change their roles at will. The topics were debatable social issues, such as the use of nuclear power, internet safety, and free trade. For OPI sessions, participants submitted video recordings of their Zoom sessions to the cloud each time. In addition, the learners were asked to attend a commercial online conversation session, with which all of them had a yearlong special contract, at least once a week with a foreign national NNS conversation partner (online intercultural interaction; OII).

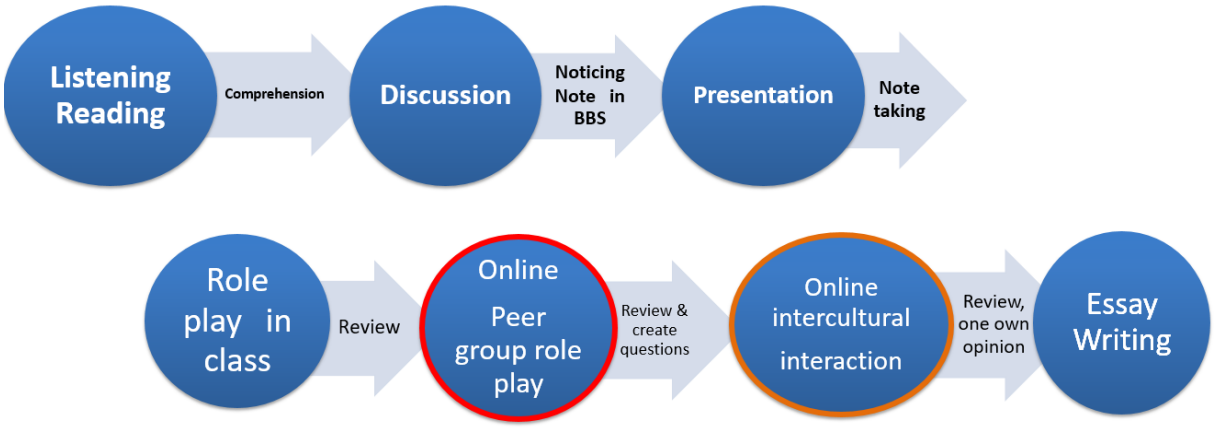


Figure 1. Active learning cycle in the special seminar: Discussion, presentation, and role play were conducted in 100-minute classes; listening/reading, online peer-group role play, and the follow-up activities were done outside of class on an individual/group basis.

2.3. Measurement

In order to measure the effects on speaking ability and attitudes toward the English language, a mixed-method research design was adopted. Speaking tests conducted by DMM, a commercial online conversation program in Japan, were administered in April, July, and December of 2021. The speaking tests scores were based on one-on-one interaction using DMM's original scheme (see Table 1). The scoring scale was from 2 (beginner) to 8 (advanced).

Table 1. Speaking test components (DMM, 2023).

Speaking test tasks	Points
(1) Reading aloud a news article or an expository passage after one minute of silent reading	5
(2) Question and answers about the text read aloud: three comprehension questions	5
(3) Picture narration: Within one minute after planning for 30-seconds	5
(4) [Basic] Making a two-minute speech on one's personal life; [Advanced] Compare two statistical graphs and tell what can be said from them after planning for one minute	5
(5) [Basic] Q&A about the speech in (4); [Advanced] Q&A about the issue described in (4)	10
Holistic level judgement: 2 (Beginner) – 5 (Intermediate) – 8 (Advanced)	

In order to examine the learners' attitudes toward EFL/EIL, the *English as an international language perception scale* (Nakamura, Lee, & Lee, 2018) was used. This questionnaire was conducted in April and December. The instrument consisted of four categories, with each of them having three to four items. Participants responded to the statements using a 5-point scale.

Interaction during the online role play task was assessed by analyzing ten videos of participants who had shown significant improvement on the speaking tests. Transcripts of the role plays were analyzed for negotiation for meaning: trigger – indicator – response – reaction to response (Varonis & Gass, 1985).

3. Results

The results of the speaking tests at three points in time indicated that a statistically significant improvement was observed with a large effect size (Table 1). Post hoc comparisons for the total test scores showed a statistically significant increase from April to December ($p < .05$).

Table 2. Results of the speaking tests.

	April		July		December				
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*	η_p^2
Q1	3.79	0.85	4.53	0.51	4.63**	0.76	16.70	.001	0.48
Q2	3.21	1.03	3.68	1.06	3.95	1.18	5.16	.074	0.20
Q3	3.89	0.74	4.32	0.75	4.53**	0.51	5.03	.012	0.22
Q4	3.63	0.76	3.63	0.68	4.00	0.82	1.69	.198	0.09
Q5	6.68	1.49	6.89	1.37	7.11	1.63	0.88	.360	0.05
Overall	21.21	3.94	23.05	2.68	24.21*	3.19	5.60	.008	0.24
Level	5.89	0.94	6.32	0.67	6.58	0.69	4.85	.014	0.21

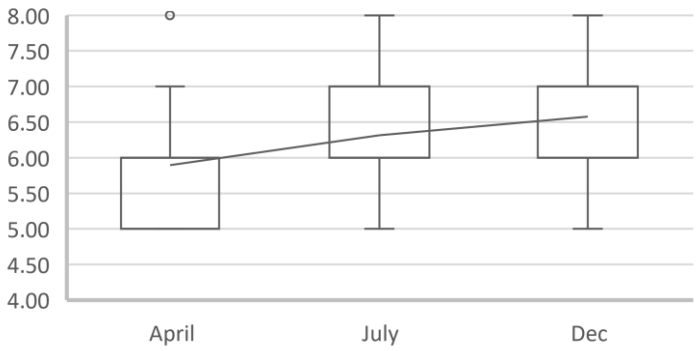


Figure 2. Box plots showing changes in speaking skills over time (levels).

As for the results of the questionnaire, a significant difference with a moderate effect size was found for scores derived from the fourth category of the instrument which examined participants' perceptions of English speakers' identities, $t(16) = 2.17$, $p = .046$, $d = .53$. No significant differences were observed for the other three categories (i.e. understanding of the current status of English; attitude toward varieties of English; and strategies for multilingual/multicultural communication).

Table 3. Descriptive statistics for the fourth category of the questionnaire.

Perception of English Speakers' Identity	Time 1 (April)		Time 2 (Dec)	
Items	*M*	*SD*	*M*	*SD*
(1) English teachers should not push me to speak like a 'native' English speaker.	3.29	0.92	3.82	1.07
(2) I don't mind if people laugh at my English accent when I speak because it is my own English.	2.82	0.88	3.24	1.03
(3) It is unnecessary to speak like American or British English speakers as long as my English is intelligible to others.	3.06	1.20	3.47	1.18

In the qualitative analysis of online peer interaction, spoken data from ten Zoom recordings, five from April and five from December, were examined. Occurrences of Negotiation for Meaning (NfM) were observed in five videos (50%). L1 use for unknown words was seen as a trigger and indicator of NfM as were other active conversational features, such as asking extended questions and commenting on opinions (Table 4).

Table 4. Instances of negotiation for meaning observed during online role play videoconferencing.

Zoom recording 1 (19'28" 26 turns) Spring
NfM 1 (5'20"-5'34") R: What kind of measure [trigger]do you think the government should take? K: Measure? What kind of measure? [indicator] R: ... Like, measure is a solution. [response] K: Like a solution [reaction to response] NfM 2 (12'43"-13'40") R: Everyone does not get vaccinated (Trigger) K: Does not get vaccinated (Indicator: repetition) R: Still young people do not get vaccinated. Still no *wakuchin.* (R: repetition & L1 translation) K: Ah, *Wakuchin*! So we should take vaccine for Olympics. (RR: confirmation & modified output)
Zoom recording 2 (31'44" 44 turns) Spring
NfM 3 (9"16"-10' 25") K: ...only pay for student who is, *nante iun daro na, yarukiga aruhito tte*? (What should I say for a person who has motivation?) [trigger & self-initiated indicator: using L1 to ask for help] Y: People who have motivation. [R: L2] K: Who have motivation and Japanese government pay education fee for students who have more motivation [RR: repetition & modified output]
Zoom recording 3 (15'49" 24 turns) Spring
NfM 4 (10'34"-10'53") Y:...It is contemporary, *janakute* (not this word). [trigger/self-initiated indicator in L1] B: Temporary. (R: co-construction) Y: Temporary. So the benefit is temporary. (RR: repetition & modified output)
NfM 5 (12'25"-14'12") Y: For example, like snowmobile stadium... (T) A: Are there snowmobiles as Olympic summer sports? Y: Sorry? (Clarification request) A: So you say stadium for special sports like snowmobile, right? (Confirmation check) Y: Yes. (Indicator) A: Is there snowmobile? Yes? (R: Clarification) B: Snowboarding, *janai*? Snowboard, snowboard. Yeah? Ski, ski. (R: Correction) Y: (L1) All right. I just wanted to mention some sports because I don't know many names of sports. (RR: mentioning the reason in the L1 for the mistake)
Zoom recording 4 (19'28" 26 turns)
Asking questions referring to previous utterances: R: Do you guys have any questions to each other? Y: Yeah, I have a question. You said that we build TPP without China, but ... what do you think about it? (Active listening and *Wh*-question)

4. Discussion

The first research question examined the effect of group-based videoconferencing in EFL on L2 speaking skills. The results indicated that the participants' L2 speaking skills improved over time, with statistically significant gains on interactive speaking tests. This finding corresponds to the results of Lenkaitis (2020) in which participants reported increased satisfaction with their L2 speaking skills after six-weeks of videoconferencing in L2 Spanish. The opportunities provided by online EFL conversation and peer interaction in the current study could have motivated the learners to use English for communicative purposes and led to increased oral communication skills.

The second research question asked about the extent to which videoconferencing in EFL influences L2 learners' attitudes toward EFL. Although scores from three sections of the questionnaire did not significantly change over the course of the research, scores related to participants' perceptions of English speakers' identities significantly increased, which supported the findings of Iino (2022). This result could indicate that the participants began to develop attitudes that were more accepting of varieties of English and more forgiving regarding their own and others' non-native-like language use. Exposure to and participation in communicative environments in which ELF is used could bring about positive attitudinal changes which promote L2 learning.

Finally, the third research question looked at qualitative aspects of online group role plays. L1 use was observed to some extent, which often facilitated comprehension. In sessions without NfM, it is possible that the participants prioritized continuing the conversation without giving negative feedback or asking for assistance. In sessions with NfM, the third-year students often showed supportive attitudes to the younger students; this support may have made the students less anxious and less hesitant to modify their output (Philp, Adams, & Iwashita, 2013).

5. Conclusions

The purpose of the current study was to examine the effects of videoconferencing in EFL. Overall, the results provided evidence for the facilitative effects of L2 interaction through videoconferencing and the communicative tasks on L2 learners' speaking skills and attitudes toward EFL. Standardized speaking tests revealed increases in L2 learners' speaking ability over time, and the learners' showed more acceptance to varieties of English to which they had little previous exposure. Furthermore, the qualitative results indicated that NNS-NNS videoconferencing can provide opportunities for learners to negotiate meaning and support each other throughout the learning process. Future studies could examine which aspects of L2 speaking skills change over time in synchronous computer-mediated environments and how learners qualitatively perceive interlocutor background variables.

Acknowledgements

This work was supported by JSPS KAKEN Grant No. 23K00731 (Principal Investigator: Atsushi Iino; Co-Investigators: Brian Wistner, Hitoshi Akutsu, Ito Takehiko, & Jay Tanaka).

References

Adams, R., & Oliver, R. (2019). *Teaching through peer interaction*. Routledge: New York.

Iino, A. (2022). Effects of incorporating online English conversation into an active learning cycle on speech improvement and cross-cultural understanding. *Kaizai Shirin* (The Hosei University Economics Review), *89*(4),161–182.

DMM Eikaiwa. (2023). Speaking tests (briefing online site). https://eikaiwa.dmm.com/app/materials/speaking-tests/7VY89NtREeiYvut8If6MhA

Hetrovicz, L. (2021). The effect of NNS–NNS and NNS–NS videoconferencing on the development of second language confidence. *Foreign Language Annals*, *54*, 1257–1277. https://doi.org/10.1111/flan.12592Lenkaitis

Lenkaitis, C. A. (2020). Technology as a mediating tool: Videoconferencing, L2 learning, and learner autonomy. *Computer Assisted Language Learning: An International Journal, 1*(33), 483–509.

Nakamura, Y., Lee, J. S. J., & Lee, K. (2018). English as an international language perception scale: Development, validation, and application. *Language, Culture and Communication, 50*, 189–208.

Philp, J., Adams, R., & Iwashita, N. (2013). *Peer interaction and second language learning*. Routledge: New York.

Varonis, E., & Gass, S. M. (1985). Miscommunication in native/non-native conversation. *Language in Society, 14*(3), 327–343.

Warner-Ault, A. (2020). Promoting intercultural learning through synchronous video exchange: A talk abroad case study. *International Journal of Computer Assisted Language Learning and Teaching, 10*(1), 1–14.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16916

Effects of text-to-speech synthesized speech on learners' presentation anxiety and self-efficacy: A comparison of two models

Takatoyo Umemoto[a], Shinnosuke Takamichi[b], Yuta Matsunaga[c], Yusuke Yoshikawa[d], Kikuko Yui[e], Kishio Sakamoto[f], Shigeo Fujiwara[g], and Yasushige Ishikawa[h]

[a]Institute for Liberal Arts and Science, Kyoto University of Foreign Studies, , t_umemoto@kufs.ac.jp; [b]The University of Tokyo, , shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp; [c]The University of Tokyo, , matsunaga-yuta339@g.ecc.u-tokyo.ac.jp; [d]Kyoto University of Foreign Studies, , y_yoshikawa@kufs.ac.jp; [e]Kyoto University of Foreign Studies, , k_yui@kufs.ac.jp; [f]Kyoto University of Foreign Studies, , k_sakamoto@kufs.ac.jp; [g]Uchida Yoko Co., Ltd., , shigeo.f@uchida.co.jp and [h]Kyoto University of Foreign Studies, , y_ishikawa@kufs.ac.jp

How to cite: Umemoto, T.; Takamichi, S.; Matsunaga, Y.; Yoshikawa, Y.; Yui, K.; Sakamoto, K.; Fujiwara, S.; Ishikawa, Y. (2023). Effects of text-to-speech synthesized speech on learners' presentation anxiety and self-efficacy: A comparison of two models. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16916

Abstract

This paper reports on the effects of two Text-To-Speech (TTS) synthesized speech models, one based on English utterances by a native English speaker and the other based on English utterances by a Japanese non-native speaker of English, on presentation anxiety and self-efficacy of Japanese English as a Foreign Language learners. We hypothesized that learners' presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared with the native English speaker model. 55 first-year university students (upper level: 33; lower level: 22) voluntarily participated in the study and were divided into experimental and control groups. To measure the participants' presentation anxiety and self-efficacy, the scale developed by Ishikawa et al. (2021) was used. A mixed-design three-factor ANOVA with the group, class level, and period as independent variables showed an interaction between class level and period. A simple main effect test indicated a significant increase in self-efficacy in the upper-level students. These results reveal that, regardless of the model used, the use of TTS-synthesized speech significantly increases the self-efficacy of the upper-level students. The paper concluded that further research on technology use and learner affect needs to be conducted.

Keywords: *presentation anxiety, self-efficacy, text-to-speech, synthesized speech.*

1. Introduction

Studies have demonstrated the usefulness of Text-To-Speech (TTS) in language-learning settings. Handley and Hamel (2005) reported positive results for TTS use and recommended the technology for English as a Foreign Language (EFL) learners' listening and speaking practice. TTS synthesis, which generates speech from text input, offered means of providing spoken language input to learners in Computer-Assisted Language Learning (CALL) environments (Handley, 2009). A trainee teacher in Turkey utilized a web-based TTS tool outside class in order to improve her English pronunciation, and it was found that her accent started being perceived as native, indicating that the online TTS tool may be effective as a self-study tool for improving the trainees' English

pronunciation (Ekşi, & Yeşilçınar, 2016). Liakin et al. (2017) examined the impact of the pedagogical use of mobile TTS on the L2 acquisition of French liaisons. The results showed that the mobile TTS technology complemented and enhanced L2 pronunciation teaching. In addition, Chiang (2019) compared traditional teacher-led dictation and dictation with TTS among EFL learners' vocabulary performance and found a significant difference. Even in an environment where English is the first language, TTS technology has been used for English language lessons. Parr (2013) conducted an eight-month survey with 28 grade five students whose first language was English, and revealed that the TTS technology promoted an inclusive reading practice that facilitated language learning for students with different reading abilities.

Regarding learners' perceptions of pedagogical TTS use, Bione et al. (2016) found that participants had positive attitudes toward it. According to Papin and Cardoso (2022), learners viewed the self-directed learning experience with TTS and the automatic speech recognition features of Google Translate as positive. Moon (2020) conducted a questionnaire survey on using self-generated listening materials based on TTS and found that most students felt that TTS-based listening materials reduced their listening anxiety and increased their confidence.

Although previous studies have investigated the usefulness of TTS in language-learning settings, no study has compared different models of TTS-synthesized speech and their effects on the emotional aspects of learner engagement. Therefore, in this study, we developed two TTS-synthesized speech models and an interface that allows learners to use them. We hypothesized that learners' presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared to when using the native English speaker model.

2. Method

2.1. Development

Two TTS-synthesized speech models were created: one based on English utterances by a native English speaker and the other based on English utterances by a Japanese non-native speaker of English. The speech of both models was synthesized by a deep learning model 'FastSpeech 2: Fast and High-Quality End-to-End Text-to-Speech' (https://arxiv.org/abs/2006.04558v8) using data in which text and speech were paired.

The interface to allow learners to use the two above-mentioned models was developed as a web-based application that runs on students' mobile devices, such as laptops, tablets, and smartphones. After logging into the application, students can listen to the TTS-synthesized speech by the following three steps: (1) Click on the 'Create New' button, and when the dialog box appears, select the document you want to convert to a TTS-synthesized speech, and then, click the 'Start' button. (2) Click the 'Convert to Audio' button at the bottom left of the screen to convert it to a TTS-synthesized speech. (3) When the conversion is completed, a playback bar appears in the lower-left corner of the screen, which can be pressed to listen to the TTS-synthesized speech (see Ishikawa et al., 2021 for more details).

2.2. Participants

Two class levels of first-year students (upper:33; lower:22; total:55) from a Japanese university volunteered to participate in this study and were randomly divided into two groups (A:26; B:29). They were divided into two classes at their university according to their scores on the Test of English for International Communication (TOEIC) for Listening and Reading. The participants' mean TOEIC scores were 537.12 (SD = 88.43) and 456.67 (SD = 92.92) in the upper and lower group, respectively.

2.3. Instrument

Ishikawa et al.'s (2021) questionnaire was used to measure the participants' presentation anxiety and self-efficacy. The questionnaire included ten items (five items each for presentation anxiety and self-efficacy), such

as “I feel anxious when I give a presentation in English” and “I am confident in my ability to give presentations in English.” Participants rated their responses on a five-point Likert scale (1 = disagree to 5 = strongly agree).

2.4. Procedure

A pre-model survey using the above-mentioned questionnaire was administered in late November 2022. The participants then practiced their presentations for approximately three weeks outside the classroom. The participants in Group A used the non-native Japanese speaker of the English model, whereas those in Group B used the native English speaker model. In mid-January 2023, a post-model survey was conducted using the same questionnaire after the participants delivered their presentations in class. A mixed-design two-factor Analysis of Variance (ANOVA) was conducted with the groups and periods as independent variables. In addition, a mixed-design three-factor ANOVA was conducted with groups, periods, and class levels as independent variables.

3. Results

Cronbach’s alpha coefficients for presentation anxiety and self-efficacy were calculated. The survey results were sufficiently reliable (presentation anxiety: pre: α = .88, post: α = .84; self-efficacy: pre: α = .79, post: α = .83).

The two-factor ANOVA results showed no differences between the groups and periods. Furthermore, no interaction effect was observed, implying that our hypothesis was invalid. However, the three-factor ANOVA showed an interaction effect between class level and period (F (1, 51) = 5.21, $p < .05$, η_p^2 = .07; see Table 1). Thus, a simple main-effect test (the Holm method) was conducted, and a significant increase in self-efficacy was found among upper-level students. This finding indicates that the self-efficacy of upper-level students increased significantly when using TTS-synthesized speech, regardless of the model used.

Table 1. Mean and standard deviation for each variable by group, class level, and period.

	Group	Class level	Pre		Post	
			Mean	*SD*	*Mean*	*SD*
Presentation anxiety	Group A	Upper	3.55	0.84	3.54	0.57
		Lower	3.42	0.74	3.38	1.33
	Group B	Upper	3.55	1.10	3.41	0.85
		Lower	3.46	0.67	3.26	0.94
Self-efficacy	Group A	Upper	2.93	0.52	2.96	0.61
		Lower	3.04	0.47	2.89	0.96
	Group B	Upper	2.55	0.58	2.89	0.54
		Lower	2.95	0.58	2.85	0.69

4. Discussion

Our study hypothesized that learners’ presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared to the native speaker model. However, this hypothesis was not supported. Nevertheless, the three-factor ANOVA and a simple main test revealed that upper-level students’ self-efficacy increased significantly when using TTS-synthesized speech, regardless of the

model used. This outcome implies that presentation practice using TTS-synthesized speech was closely associated with participants' English proficiency. By practicing their presentations using TTS-synthesized speech, the participants in the upper-level group felt that their presentations in class would proceed smoothly. However, the participants in the lower-level group did not have sufficient English language skills to attain such confidence. This presentation practice was an out-of-classroom self-study (participants decided whether they wanted to practice their presentations). Therefore, participants in the lower-level group might have been less motivated to engage in presentation practice than those in the upper-level group.

The study's limitations are as follows: 1) the intervention used a one-shot, three-week design and 2) the presentation practice using TTS-synthesized speech was participant-directed.

Based on the results and limitations described above, there are two directions for future research. The intervention period using TTS-synthesized speech should be expanded from three weeks to one semester and incorporated into a course designed to help students improve their speaking skills and give presentations in English, offering them opportunities to practice their presentations with their instructors. Moreover, TTS-synthesized speech factors that affect learners' presentation anxiety, self-efficacy, and motivation should be assessed qualitatively, similar to Teng and Wang's (2021) research on Chinese EFL learners.

Acknowledgements

While conducting the research described in this paper, we were deeply saddened by the sudden loss of our co-researcher, Dr. Masao Aikawa, on December 18, 2022. He was a Professor at the Department of British and American Studies and Vice President at Kyoto University of Foreign Studies, Japan. He contributed significantly to TESOL and foreign language education policy research in Asia. We offer sincere condolences to his family.

This research was supported by the Intra-University Research of Kyoto University of Foreign Studies, Japan, and Grants-in-Aid for Scientific Research provided by the Japan Society for the Promotion of Science (#18K00763). The data presented, statements made, and views expressed are solely the responsibility of the authors.

References

Bione, T., Grimshaw, J., & Cardoso, W. (2016). An evaluation of text-to-speech synthesizers in the foreign language classroom: Learners' perceptions. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture—Short papers from EUROCALL 2016* (pp. 50–54). Research-publishing.net. https://doi.org/10.14705/ rpnet.2016.eurocall2016.537

Chiang, H.-H. (2019). A comparison between teacher-led and online text-to-speech dictation for students' vocabulary performance. *English Language Teaching*, *12*(3), 77–93. https://doi.org/10.5539/elt.v12n3p77

Ekşi, G. Y., & Yeşilçınar, S. (2016). An Investigation of the effectiveness of online text-to-speech tools in improving EFL teacher trainees' pronunciation. *English Language Teaching*, *9*(2), 205–214. https://www.ccsenet.org/journal/index.php/elt/article/view/56606

Handley, Z. (2009). Is Text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, *51*(10), 906–919. https://doi.org/10.1016/j.specom.2008.12.004

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

Ishikawa, Y., Takamichi, S., Umemoto, T., Aikawa, M., Sakamoto, K., Yui, K., Fujiwara, S., Suto, A., & Nishiyama, K. (2021). Japanese EFL learners' speaking practice utilizing text-to-speech technology within a team-based flipped learning framework. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies: New challenges and learning experiences* (pp. 283–291). Springer. https://doi.org/10.1007/978-3-030-77889-7_19

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning*, *30*(3-4), 325–342. https://doi.org/10.1080/09588221.2017.1312463

Moon, D. (2020). Learner-generated digital listening materials using text-to-speech for self-directed listening practice. *International Journal of Internet, Broadcasting and Communication*, *12*(4) 148–155. http://dx.doi.org/10.7236/IJIBC.2020.12.4.148

Papin, K., & Cardoso, W. (2022). Pronunciation practice in Google Translate: Focus on French liaison. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds.), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 322–327). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1478

Parr, M. (2013). Text-to-speech technology as inclusive reading practice: Changing perspectives, overcoming barriers. LEARNing Landscapes Journal, *6*(2), 303–322. https://doi.org/10.36510/learnland.v6i2.618

Teng, Y., & Wang, X. (2021). The effect of two educational technology tools on student engagement in Chinese EFL courses. *International Journal of Educational Technology in Higher Education*, *18*, 27. https://doi.org/10.1186/s41239-021-00263-0

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16984

Enhancing pronunciation instruction for non-francophone immigrants in Quebec: a technology-assisted proposal

Geneviève Bibeau[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, ge_bibea@live.concordia.ca and [b]Department of Education, Concordia University, walcir.cardoso@concordia.ca

How to cite: Bibeau, G.; Cardoso, W. (2023). Enhancing pronunciation instruction for non-francophone immigrants in Quebec: a technology-assisted proposal. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16984

Abstract

This paper has two objectives. In Part 1, we report the findings of a mixed-method study that examines the pronunciation needs of non-francophone immigrants in Quebec after they complete the Program for the Linguistic Integration of Immigrants (known as "francization"), a language learning initiative to equip non-French-speaking immigrants with essential French skills. The findings indicate a noticeable disparity between the instruction provided to learners in the program and their practical requirements in real-life situations, and a strong need by the participants to improve their pronunciation autonomously post francization. Part 2 of the study addresses the pedagogical implications of these findings, in which we address our participants' needs with a set of technology-enhanced pedagogical recommendations for blended and autonomous learning.

Keywords: *L2 French, automatic speech recognition, text-to-speech synthesis, intelligent personal assistants, L2 pronunciation.*

1. Introduction

Acquiring intelligible pronunciation in a second language (L2) can be a daunting task for many adult learners (DeKeyser, 2012). For immigrants, this challenge extends beyond the language classroom and can have a significant impact on their social, economic, and cultural integration (Simpson & Whiteside, 2015), given that intelligible pronunciation is a vital component of effective oral communication (Levis, 2020). In Quebec, the Program for the Linguistic Integration of Immigrants (PILI, also known as *francization*) strives to provide non-French-speaking immigrants with the means to acquire the language skills necessary to obtain employment or pursue education in the province (Ministère de l'Immigration, de la Francisation et de l'Intégration, 2022). Despite these efforts, the majority of French learners do not reach level 8, which is considered the threshold for linguistic autonomy (Vérificateur général du Québec, 2017).

To address the issue at hand, this project has two objectives. In Part 1, we report the findings of a mixed-method study that examines the pronunciation needs of non-francophone immigrants in Quebec. Considering the time constraints in traditional classroom settings, which often limit the attention given to pronunciation, and the necessity of continuing to learn it in real life, we aimed to identify the types of pronunciation instruction that the

participants received during their francization and to determine their current (post-francization) pronunciation needs. We asked the following research questions:

(1) What pronunciation instructions have the adult immigrants received during the program?
(2) Are these instructions sufficient for real-life needs?
(3) Do they enable them to continue to improve the intelligibility of their pronunciation autonomously?

Based on the insights gained from our research, including the fact that the majority of Quebec immigrants rely on smartphones as their only technological device, Part 2 addresses the pedagogical needs of these students with a set of technology-enhanced pedagogical recommendations for blended and autonomous learning using smartphones. To facilitate interaction (Chapelle, 2001) and motivate learners to practice their aural and oral skills (e.g. Cardoso, 2022), these activities follow cognitive interactionist approaches to CALL (Chapelle, 2001). They harness technologies that offer learners human-machine interactions, such as text-to-speech synthesizers (following Liakin et al., 2015), automatic speech recognition (Liakin et al., 2017), and intelligent personal assistants like Amazon Alexa (Moussalli & Cardoso, 2020).

2. Method

Part 1 collected data from 68 participants (adult immigrants, French L2 learners) using a mixed-method approach that included both qualitative and quantitative data collection and analysis techniques. A customized survey was used (following insights from Celce-Murcia et al., 2010 and Pawlak, 2010). Among its 50 five-item Likert-scale questions, 26 items assessed the frequency with which Pronunciation Instruction (PI) was received during the program (from "never" to "very often"). To capture the different types of instruction, this section was divided into three subgroups: the first covers the elements of Phonetics Commonly Taught (PCT). These are classroom instructions that encompass both segmental and suprasegmental features, taking into account three general stages of phonological development by Celce-Murcia et al. (2010): (1) sound awareness (e.g. drawing attention to a particular sound); (2) perception (e.g. through listening to minimal pairs); and (3) controlled or spontaneous oral production (e.g. reading aloud). The second subgroup pertains to Corrective Feedback (CF) given by the teacher, while the third comprises a set of Techniques, Tools, and Strategies promoting Autonomy (TTSA), based on the four categories defined by Pawlak (2010): cognitive, metacognitive, social, and affective.

Regarding the participants' real-life pronunciation needs (NEEDS) and their ability to meet them on their own (AUTONOMY) after the program, ten items measured their attitudes (from "strongly disagree" to "strongly agree") in addition to four open-ended questions. Finally, semi-structured interviews were conducted with 13 randomly selected participants to gain a deeper insight into their pronunciation needs.

The internal validity of the survey was tested in an upstream pilot, after which the internal consistency of the survey items was also assessed for each construct, showing a Cronbach's alpha greater than 0.7. The Likert scale nominal values for frequency and attitude were analyzed using descriptive statistics including Means (M) and Standard Deviations (SD). Correlations between the three different types of pronunciation instruction (PCT, CF, TTSA) and the two constructs, NEEDS and AUTONOMY, were calculated to determine the extent to which the instructions meet real-life needs and support the autonomous pursuit of learning. Finally, the interview data were analyzed qualitatively following Saldaña's (2021) cyclic methods, based on the themes described above.

3. Results

To answer the first research question, survey results indicate that participants receive a variety of pronunciation instructions in the *francization.* Overall, most PI elements were used often (16/26). However, when looking at specific subgroups, as illustrated in Table 1, PCT and CF were used often (above 3/5), while TTSA was used infrequently (2.88/5). The interview data confirmed that lessons are confined to popular features of French phonology (e.g. *liaison*, *enchaînements*), with little or no reference to how learners can continue to improve their pronunciation on their own, such as through the use of technology.

Table 1. Pronunciation instructions received during PILI by subgroups

Pronunciation Instructions (PI)	M /5	SD
Elements of Phonetics Commonly Taught (PCT)	3.29	1.03
Corrective Feedback (CF)	3.34	0.95
Techniques, Tools and Strategies for Autonomy (TTSA)	2.88	1.13

To answer our second research question, which asked whether the instructions received during *francization* were sufficient for real-life needs, survey results showed high scores for students' needs being met (NEEDS: M=3.62, SD=1.04). A positive correlation was observed between the three types of instruction and NEEDS satisfaction (Table 2). However, most students still expressed wanting to improve pronunciation (3.54, SD=1.07), regardless of the amount of instruction received.

To answer the third research question (whether instruction promotes autonomous learning), correlation analyses show that all three types of pronunciation instruction are positively correlated with AUTONOMY (Table 2). The results indicate that TTSA is the most positively correlated of the three types of instruction with NEED satisfaction and AUTONOMY.

Table 2. Correlations between pronunciation instruction (PCT, CF, TTSA) and NEEDS and AUTONOMY

	NEEDS	AUTONOMY
Elements of Phonetics Commonly Taught **(**PCT)	,452** $p < ,001$	,395** $p < ,001$
Corrective Feedback (CF)	,363** $p = ,005$	,360** $p = ,003$
Techniques, Tools and Strategies for Autonomy (TTSA)	,566** $p < ,001$	,559** $p < ,001$

**. Correlation is significant at the 0.01 level (two-tailed).

Finally, the interviews revealed a missing ingredient in the participants' learning experience. Despite survey responses suggesting they knew 'what' to do for autonomous pronunciation learning (M=4.29, SD=0.95), many did not know 'how' to do it. Survey results showing low scores for TTSA (M=2.88; Table 1) supports this evidence of a lack of strategy for autonomous pronunciation learning in the *francization*.

4. Discussion and pedagogical recommendations

Part 1 aimed to determine whether there is an alignment between the pronunciation instruction offered in the *francization* for adult immigrants and their real-life needs, including the need to continue developing their pronunciation skills independently. Despite the wide variety of instruction received, learners wish to continue their learning and improve their pronunciation outside the classroom, but they do not know how to do it. This desire echoes many studies (Derwing & Rossiter, 2002), which point to the limited opportunities for classroom interaction and personalized feedback, as well as the disparities between classroom instruction and real-life contexts (Beaulieu et al., 2021). Since learner autonomy is important for acquiring L2 pronunciation (Pawlak, 2010), TTSA should be given greater prominence, especially as it shows the strongest correlation with the constructs of NEEDS and AUTONOMY concerning the ability to continue pronunciation learning.

In response to the needs identified (i.e. that students need strategies to pursue autonomous pronunciation learning), Part 2 presents technology-assisted pedagogical recommendations likely to provide concrete ways to achieve these needs. These recommendations leverage smartphones, a widely used device that naturally lends itself to autonomous pronunciation practice.

Pronunciation instruction, as recommended by Celce-Murcia et al. (2010) and Chapelle (2001), should promote access to input (e.g. listening), opportunities for output practice (e.g. speaking, pronouncing), and input/output interaction. To meet these objectives, we propose the use of three technologies accessible via smartphones: Text-To-Speech synthesis (TTS), Automatic Speech Recognition (ASR), and Intelligent Personal Assistants (IPA).

TTS is a technology that converts written text into spoken speech. This allows learners to develop sound awareness and perception by listening to individual words or full texts read aloud. TTS facilitates access to pronunciation models and can improve listening skills (Liakin et al., 2017). ASR, on the other hand, enables learners to practice their pronunciation and speaking skills. It provides real-time written feedback on their speech, helping them identify and improve specific pronunciation difficulties. Finally, IPAs facilitate communicative practice. Through interacting with IPAs, learners can ask questions and receive spoken responses, promoting low-stress and dynamic input/output interactions.

Teachers play a key role in implementing these recommendations in all stages of pronunciation learning (Celce-Murcia et al., 2010), from developing sound awareness (using TTS) to oral production (ASR) and communicative practice (IPA). By associating these stages with TTS and ASR (e.g. as found in Google Translate) and with the participants' smartphones' IPA, teachers can equip their students for effective autonomous learning. As illustrated in Figure 1, using these speech technologies in a sequenced manner can give learners concrete strategies to improve aspects of their pronunciation autonomously. For instance, they can listen to models (TTS), pronounce, repeat, and adjust speech (ASR), and then practice conversations (IPA). This builds proficiency in the pronunciation features from class while providing strategies for continued autonomous learning.

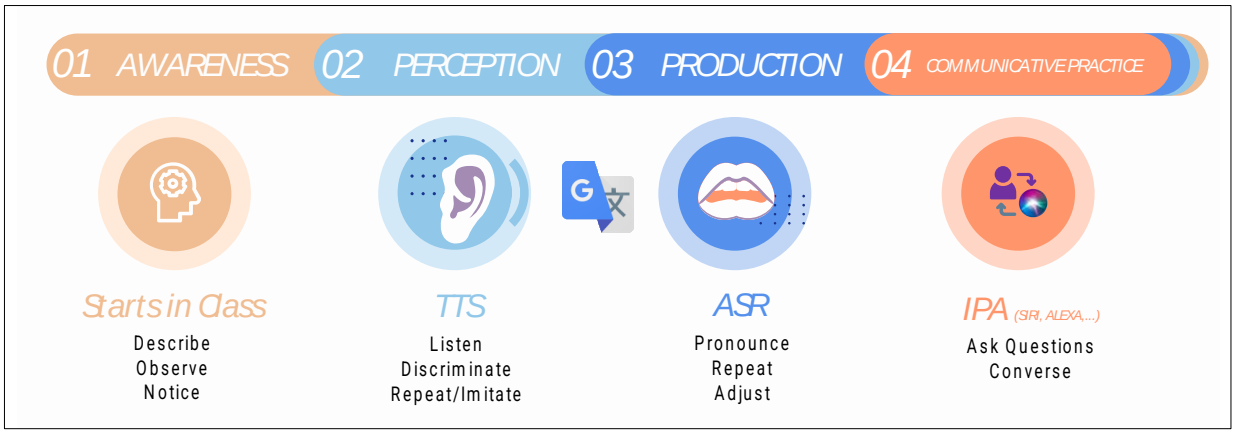


Figure 1. Pronunciation stages, smartphone affordances, and possible actions.

For example, after learning about French nasal vowels in class, students could practice aural perception through a "same or different" minimal pair activity using TTS audio. The teacher could provide minimal pair words containing nasal vowels for students to distinguish, allowing them to develop perceptual skills for the pronunciation feature covered in class, as illustrated in Figure 2.

	Minimal pairs: nasal vowels		Same (S) or Different (D)
1.	Temps.	Thym.	
2.	Tas.	Temps.	
3.	Tant.	Temps.	
4.	Ton.	Taon.	
5.	Temps.	Taon.	
6.	Vent.	Vend.	
7.	Vin.	Vent.	
8.	Vas.	Vent.	
9.	Vont.	Vent.	
10.	Brin.	Brun.	
11.	Plan.	Plein.	
12.	Pont.	Paon.	
13.	Banc.	Ban.	
14.	Pain.	Peint.	
15.	Daim.	Dont.	

Figure 2. Same or different? Developing perception of nasal vowels with TTS.

Students could then practice producing nasal vowels by speaking to ASR and checking its output (orthography) for accuracy (e.g. if they say "vin" but the transcription reads "vent", that will serve as an indication that their pronunciation is inaccurate). Finally, they can use a set of words containing nasal vowels and engage in conversation with an IPA to practice them (e.g. words such as *temps*, *temperature*, or *boire du vin* could be elicited when discussing one's summer). This allows communicative practice in producing the target feature/s in a low-stress conversation.

Familiarizing students with these three speech technologies during *francization* holds the potential for them to develop strategies that extend beyond the classroom, empowering them to meet their real-life pronunciation needs on their own.

5. Conclusions

This study had two objectives, conceptualized as Parts 1 and 2. Part 1 reported the findings of a mixed-method study that explored the pronunciation needs of non-francophone immigrants in Quebec after *francization*. The results highlighted the need for the teaching of strategies to sustain pronunciation learning autonomously. Part 2, on the other hand, addressed the pedagogical needs of these students by proposing technology-enhanced recommendations for blended (during *francization*) and autonomous learning (after *francization*). These recommendations leverage the capabilities of three smartphone-available technologies: TTS, ASR, and IPA. However, as is the case with any research, there are some limitations to acknowledge for future research. These include a small sample size, the shortcomings of these technologies (e.g. recognition errors), and no empirical data to support the effectiveness of the proposed recommendations.

Given the widespread smartphone ownership in Canada (95% under age 50; Pew Research Center, 2022), pronunciation strategies using built-in tools hold potential, as they can increase learning efficiency and autonomy. These tools should be taught early in *francization* programs to empower learners and address critical communication needs outside class and in their future language learning endeavours.

Acknowledgements

We would like to thank the 68 participants who generously volunteered their time, making this study possible.

References

Beaulieu, S., French, L. M., Bejarano, J., & Reinke, K. (2021). Cours de français langue seconde pour personnes immigrantes à Québec : portrait des habiletés orales en fin de parcours. *Canadian Journal of Applied Linguistics, 24*(3), 1-29. https://doi.org/10.37213/cjal.2021.31120

Cardoso, W. (2022). Technology for speaking development. In T.M. Derwing, M.J. Munro et R.I. Thomson (Eds), *The Routledge Handbook of Second Language Acquisition and Speaking* (p. 299-313). Routledge

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: Reference for teachers of English to speakers of other languages.* Cambridge University Press.

DeKeyser, R. (2012). Age effects in second language learning. In S. M. Gass & A. Mackey (Eds) *The routledge handbook of second language acquisition* (p.442-460). Routledge.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*(2), 155-166. https://doi.org/10.1016/S0346-251X(02)00012-X

Levis, J. M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310-328. https://doi.org/10.1075/jslp.20050.lev

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

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning, 30*(3-4), 325-342. https://doi.org/10.1080/09588221.2017.1356086

Ministère de l'Immigration, de la Francisation et de l'Intégration (2022). *Programme d'intégration linguistique pour les immigrants, 2022-2023.* https://cdn-contenu.quebec.ca/cdn-contenu/francisation/MIFI/formulaires/NOR_PILI_2022-2023.pdf

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

Pawlak, M. (2010). Designing and piloting a tool for the measurement of the use of pronunciation learning strategies. *Research in Language, 8*, 189-202. https://doi.org/10.2478/v10015-010-0005-6

Pew Research Center (2022, December 6). *Internet, smartphone and social media use.* Pew Research Center. https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/

Saldaña, J. (2021). The coding manual for qualitative researchers (4e éd.). SAGE.

Simpson, J., & Whiteside, A. (2015). *Adult Language Education and Migration: Challenging agendas in policy and practice.* Taylor et Francis.

Vérificateur général du Québec. (2017, November 23). *Rapport du Vérificateur général du Québec à l'Assemblée nationale pour l'année 2017-2018, Automne 2017, Faits saillants.* https://www.vgq.qc.ca/Fichiers/Publications/rapport-annuel/2017-2018-Automne/fr_Rapport2017-2018-AUTOMNE-Faits-saillants.pdf

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16957

Enhancing short academic presentations through extended independent practice using VoiceThread

Heejin Chang[a] and Scott Windeatt[b]

[a]UniSQ College, University of Southern Queensland, Australia, heejin.chang@unisq.edu.au and [b]School of Education, Communication and Language Science, Newcastle University, UK, scott.windeatt@ncl.ac.uk

How to cite: Chang, H.; Windeatt, S. (2023). Enhancing short academic presentation through extended independent practice using VoiceThread. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16957

Abstract

This study examines the potential of VoiceThread (VT) as a learning tool to enhance learners' performance and confidence while facilitating ongoing independent practice in a 10-week English for Academic Purposes (EAP) programme in an Australian university. Students carried out weekly academic presentations on topics of personal interest using VT for recording and rehearsal. Their recordings were shared for peer review and teacher feedback. The recordings were analysed in relation to a number of features to measure utterance fluency, including speed, repairs and pauses. Questionnaires were used to investigate students' perceptions of the value of the task, of their performance on the task, and of changes in their performance and confidence over time. Interviews were conducted to investigate the potential of recording and rehearsal for long-term independent practice. In total, 34 students participated in 10-week courses, 22 students using the VT software, and 12 students in a comparison class following the same curriculum, but without the VT software. The findings suggest that regular VT tasks, over time, help students develop their spoken English performance and confidence. Such voice-recording activities hold promise for improving oral performance by encouraging independent practice in oral activities, effectively addressing the time limitations, constraints practice opportunities, and feedback challenges inherent in classroom-based courses.

Keywords: *speaking, voice recording, fluency, academic presentation, independence.*

1. Introduction

University students in many subjects are required to deliver presentations as part of their academic work, and these often form part of their assessed coursework (Grez, Valcke, & Roozen, 2009; Kerby & Romine, 2009). Helping students develop the skills required to effectively plan and deliver well-structured presentations is therefore an important element in many English for Academic Purposes (EAP) courses (Ferris, 1998). The current study was intended to tackle problems identified in the literature on oral skills (e.g. Ferris, 1998; Morita, 2000; Zareva, 2009, 2016), and by the teachers on an EAP course in an Australian university.

The main issues which the teacher identified as limiting progress were lack of time for students to practise presentations in class, or to learn from and discuss each other's presentations, and limited time for teacher feedback. Providing students with the means to rehearse and refine their presentations in their own time

using voice recording software was seen as a possible solution to this problem, and part of the course was redesigned to incorporate *VoiceThread* software (http://voicethread.com/), a cloud-based video or audio discussion application available through the university Learning Management System – *StudyDesk*. This would allow students to revise their performance until they were happy with their recordings but would also provide opportunities for feedback from the teacher, and from other students.

2. Literature Review

2.1 Academic presentations

Studies of academic presentations in L2 language learning focus on grammatical, lexical and discourse features (Swales, 2004; Zareva, 2016), or on different approaches between L1 and L2 students. The study from Zareva (2009) suggested that L1 students tend to interpret academic presentations as an opportunity to present information in an informal way, but also to interact with the audience. L2 students, however, tend to take a more formal approach focusing more on presenting information and taking less account of the reactions of their audience.

The findings of a training programme with university students in Hong Kong (Bankowski, 2010) suggest students develop analytical and thinking skills relevant to organisation, content and delivery of a presentation, but have difficulty in selecting appropriate topics. Radzuan and Kaur (2011) suggest students lack basic presentation skills and confidence in delivering a presentation, and Stapa, Murad, and Ahmad (2014) highlight the role of anxiety based on low self-perceived language proficiency in delivering effective presentations. Siddons (2008, pp. 1-2) suggests the need for academic presentations to take account of the audience, the presenter, and the presentation itself.

2.2 Fluency

The concept of fluency in speaking is not well understood and is difficult to define. Lennon (1990) defines fluency in both a broad (overall speaking proficiency) and a narrow (smoothness and ease of oral linguistic delivery as measured in terms of rate of speech) sense, and Segalowitz (2010) suggests three aspects to fluency: cognitive (the speed and manner of the underlying mechanics of speech production); perceived (the particular reaction from listeners to the cognitive fluency of the speaker); and utterance fluency (the measurable aspects of speech fluency which reflect the cognitive fluency underlying speech production). Speech rate (i.e., the number of syllables per minute, including pause time) and mean length of run (i.e., the mean number of syllables between two silent pauses) are strongly associated with the development of L2 oral fluency (Cucchiarini et al., 2002; Kormos & Denes, 2004) and with perceived fluency (Kormos & Denes, 2004; Suzuki & Kormos, 2023).

2.3 Voice-recording and VoiceThread

Many studies (e.g. Volle, 2005, Kay, 2012), have investigated techniques such as video- and audio-recording, podcasts and video-conferencing (VC) with L2 learners. Tecedor and Campos (2019) investigated voice-recording (VR), VC and face-to-face (F2F) techniques in interpersonal and presentational tasks. VC proved most effective in promoting fluency in presentational tasks, while F2F and VC tasks were equally effective in promoting complexity. The activities were, however, carried out in-class, rather than independently.

VoiceThread (VT) (https://voicethread.com/) is a cloud-based video or audio discussion application linked to a Learning Management System (LMS) such as Moodle. It allows students to plan, record, review and re-record their presentation, a process which can facilitate learners' reflection and foster the development of metacognitive strategies for monitoring their progress in oral presentations (Richardson, 2006), and provide potential for improving attitudes towards reducing language anxiety, and for increasing L2 motivation and self-confidence (Kern, 2014; Brunvand & Byrd, 2011).

To our knowledge, no research has been carried out on utterance fluency in L2 learners' academic presentations using recording applications for extended independent practice. The aim of this study was therefore to use VT

to investigate its effect on learner performance (fluency), confidence in academic presentations, and its implications for longer term independent practice.

3. Research questions

1. What changes in speaking fluency take place in student performance on an extended speaking task over time? Are the results the same for both the Experimental and Comparison Groups?

2. What are students' perceptions of the value of the recording task in terms of changes in: a. their performance? and b. their confidence in extended speaking tasks?

3. Are the students applying similar independent practice techniques beyond the current course?

4. Methodology

4.1 Course and participants

The participants were studying for 12 hours per week on ten-week EAP courses, with a short academic presentation as a final assessment. Their level on entry to the course was overall IELTS 5.0. They had lived in Australia for varying periods, spanning from one week to six years. A quasi-experimental approach was adopted, with 22 students in total forming an Experimental group (EG) in three courses using *VoiceThread* to rehearse their presentation in their own time, and receiving individual, face-to-face, weekly teacher feedback using criteria adapted from IELTS. A Comparison group (CG) consisting of 12 students, followed the same curriculum, but without using *VoiceThread*. The 34 students were from a variety of nationalities (i.e., Afghanistan (2), China (4), Taiwan (1), Congo (2), Syria (5), Iraq (14), South Korea (3), Sudan (1), Uganda (1), and Venezuela (1)).

4.2 Data collection and analysis

Utterance fluency was analysed using automatic transcription, *Otter* (http://otter.ai), and histogram software, *Praat*, (httpy://www.fon.hum.uva.nl/praat/) to identify speed, pause and repair features. Student perceptions of fluency and the value of the *VoiceThread* tasks were investigated using weekly surveys and an end-of-course questionnaire (see Chang & Windeatt, 2021, pp. 16 and 21). A follow-up semi-structured interview lasting 20-30 minutes was conducted with 10 students three to six months after the course (when most of the participants had moved on to the next level of the EAP programme, EAP2). During this interview, they were asked whether they continued to use any methods they had practiced in the course, and how and why they used them.

5. Results and discussion

R1. What changes in fluency take place in student performance on an extended speaking task over time? Are the results the same for both groups?

Mean scores were compared for the EG and CG presentations in weeks 1 and again in week 10 using a t-test, and for the EG a t-test was used to compare changes in scores from week 1 to week 10.

In the week 1 self-introduction presentation task, there was no significant difference between the groups. In the week 10 academic presentations significant differences were identified in speed ($p<.05$) and mean length of silent pauses ($p<0.1$). Within the EG there was a significant difference in mean length of silent pauses ($p<.05$) for the two presentations (week 1 and week 10), i.e. the length of silent pauses (one measure of fluency) were longer in week 10. In a survey using criteria adapted from IELTS (Chang & Windeatt, 2021, Supplementary Appendix 1), significant changes were noted in self-perceived fluency in the EG between week 1 and week 10 in four areas; 1) fluency and coherence, 2) lexical resource, 3) pronunciation, and 4) grammatical range and accuracy: significant differences in the areas of fluency and coherence ($p<.05$), and lexical resource ($p<.01$).

R2. What are students' perceptions of the value of the recording task in terms of changes in: a. their performance? and b. their confidence in extended speaking tasks?

Participants in the EG agreed or strongly agreed that recording activities helped them improve their short presentations, public-speaking skills, pronunciation, and vocabulary, and organise their ideas. They felt their fluency and grammar improved and noted that the recording and listening features of VT helped them identify their weaknesses and strengths and monitor their progress. They also agreed that they learned from other students' recordings, agreed or strongly agreed that confidence in their ability to improve their English independently had improved, and they were more relaxed when speaking to a group of people in English (though some participants, despite claiming to have grown in confidence, disagreed or strongly disagreed that they were relaxed or could speak English easily) (see Table 1).

Table 1. End-of-course questionnaire: descriptive statistics (Chang & Windeatt, 2021, p. 16)

Experimental group			
Statement		Mean*	SD
Recording activities help me to improve my	1. short presentations.	4.6	0.5
	2. speaking fluency.	4.2	0.6
	3. public-speaking skills.	4.5	0.5
	4. pronunciation.	4.3	0.7
	5. grammar.	4.1	0.5
	6. vocabulary.	4.8	0.4
	7. ideas.	4.4	0.5
Recording activities help me to	8. learn from listening to other students' recording.	4.1	0.7
	9. monitor my progress.	4.3	0.6
	10. identify my weaknesses in speaking.	4.6	0.6
	11. identify my strengths in speaking.	3.8	1.0
18. I have developed my confidence to improve English independently from the recording activities.		4.4	0.6
19. I want to keep doing the recording activities in the future.		3.7	0.6

comparison group vs. experimental group						
Statement	Mean*		SD		*df*	*p*
	CG	EG	CG	EG		
12. When I speak English, I feel cheerful.	4.0	4.1	0.8	0.5	1	0.4161
13. I am relaxed when speaking in English.	2.6	3.6	1.1	1.2	1	0.0316*
14. I can discuss and give my opinion in English.	4.0	4.4	0.8	0.5	1	0.0661
15. I will speak to a group of people in English.	3.8	4.4	0.5	0.5	1	0.0157*
16. I can be interviewed in English.	3.3	3.7	0.7	0.8	1	0.0729
17. I can speak English easily.	2.8	3.4	0.9	1.2	1	0.0782

Findings of a positive relationship between voice recording and speaking skills in previous research (Kay, 2012; Volle, 2005) were confirmed in our study, with most participants perceiving as positive the overall development of their language skills, content knowledge, digital literacy and developing autonomy. They also valued regular individual teacher feedback, which encouraged them to continue regular practice with the recording tasks.

Weekly surveys revealed a similar pattern throughout the course, with students rehearsing, then recording and listening to their recordings before uploading them to Study Desk. They accessed others' work to look for ideas about content and delivery, as part of their preparation for recording. If they found problems in their presentation (i.e. speech flow, intonation, and, less commonly, grammatical mistakes), they recorded their presentation again until they were satisfied. Being able to rehearse and listen to their own recordings as often as possible before uploading a final version appeared to help students' language skills.

R3. Are the students applying similar independent practice techniques beyond the current course?

In the end-of-course survey, 15 participants agreed or strongly agreed that they wanted to carry on using a recording activity after the course. In a post course interview with 10 participants three to six months after the course, eight students claimed that they had continued using recording to improve their speaking skills in the EAP2 course, though using mobile phone applications, which they found more convenient and user-friendly than *VoiceThread.* They focused on using techniques in the EAP2 course which they had learned on EAP1. They also adapted the recording activities to suit their own needs, for example by recording words, reading aloud, and texting voice messages.

6. Conclusions

The aim of this study was to use recording software – *Voicethread* - as a tool to enhance learners' performance on, and confidence in, extended speaking tasks in the form of academic presentations, to investigate aspects of fluency in their performance of those tasks, and to provide the basis for longer term independent practice. Our data suggest, for the experimental group, significant improvements in fluency and coherence, and in lexical resource, between weeks 1 and 10, but no significant perceived changes in pronunciation, or grammatical range and accuracy. The results for speed, pauses and repairs, show a significant change only in the mean length of silent pauses between weeks 1 and 10. The experimental group therefore appeared to improve the fluency and coherence of their presentation, used a wider range of vocabulary, but, although their rate of speech was much faster, and their silent pauses much shorter, than for the comparison group in week 10, their silent pauses were longer than in week 1. While there does, therefore, appear to be some relationship between the length of silent pause and speed of delivery, the relationship between pauses and other measures of performance such as fluency remains unclear.

Most students in the Experimental Group were positive about the benefits to be derived from the use of *VoiceThread*, including opportunities for the rehearsal, recording, re-recording, self-reflection, teacher feedback and access to peers' recordings. This was confirmed in post-course interviews with a sample of participants, who confirmed that they had continued to make use of the strategies and techniques they had developed in their EAP course, albeit with voice-recording software. This study provides sufficient evidence to justify further investigation into the potential impact of voice-recording activities, particularly on aspects of fluency, including the location, purpose and effect of pauses.

Notes

This paper summarised an article published in 2021:

Chang, H. & Windeatt, S. (2021). Using *VoiceThread* for extended independent practice in giving short academic presentations. *Computer Assisted Language Learning*, https://doi.org/10.1080/09588221.2021.2003407

References

Bankowski, E. (2010). Developing skills for effective academic presentations in EAP. *International Journal of Teaching and Learning in Higher Education, 22(2),* 187-196.

Brunvand, S., & Byrd, S. (2011). Using VoiceThread to promote learning engagement and success for all students. *TEACHING Exceptional Children,* 43(4), 28-37. https://doi.org/10.1177/004005991104300403

Chang, H. & Windeatt, S. (2021). Using *VoiceThread* for extended independent practice in giving short academic presentations. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.2003407

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustica Society of America, 107(2),* 989-999. https://doi. org/10.1121/1.428279

Grez, L. D., Valcke, M., & Roozen, I. (2009). The impact of an innovative instruction- al intervention on the acquisition of oral presentation skills in higher education. *Computers & Education, 53*(1), 112-120. https://doi.orgl10.1016/j.compedu.2009.01.005

Ferris, D. (1998). Students' views of academic aural/oral skills: A comparative needs analysis. *TESOL Quarterly, 32*(2), 289-318. https://doi.org/10.2307/3587585

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior, 28*(3}, 820-831. https://doi.org/10.1016/j.chb.2012.01.011

Kerby, D., & Romine, J. (2009). Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business, 85*(3), 172-179. https://doi.org/10.1080/08832320903252389

Kern, R. (2014). Technology as Pharmakon: The promise and perils of the Internet for foreign language education. *The Modern Language Journal, 98*(1), 340-357. https://doi.org/10.1111/j.1540-4781.2014.12065

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145-164. https://doi.org/10.1016/j. system.2004.01.001

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40(3),* 387-417. https:/1doi.org/10.1111/j.1467-1770.1990.tb00669.x

Morita, N. (2000). Discourse socialization through oral classroom activities in a TESL graduate program. *TESOL Quarterly, 34*(2), 279-310. https://doi.org/10.2307/3587953

Radzuan, N., & Kaur, S. (2011). Technical oral presentations in English: Qualitative analysis of Malaysian engineering undergraduates' sources of anxiety. *Procedia - Social and Behavioral Sciences, 29*, 1436-1445. https://doi.org/10.1016/j.sbspro.2011.11.383

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.

Siddons, S. (2008). *The complete presentation skills handbook: How to understand and reach your audience for maximum impact and success.* London: Kogan Page.

Stapa, M., Murad, N. A., & Ahmad, N. (2014). Engineering technical oral presentation: Voices of the stakeholder. *Procedia - Social and Behavioral Sciences, 118(*19), 463-467. https://doi.org/10.1016/j.sbspro.2014.02.063

Suzuki, S., & Kormos, J. (2023). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition, 45*(1), 38-64.

Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Tecedor, M., & Campos-Dintrans, G. (2019). Developing oral communication in Spanish lower-level courses: The case of voice recording and videoconferencing activities. *ReCALL, 31*(2), 116-134. https://doi.org/10.1017/S0958344018000083

Volle, L. (2005). Analyzing oral skills in voice e-mail and online interviews. *Language Learning and Technology, 9*(3), 146-163.

Zareva, A. (2009). Informational packaging, level of formality, and the use of circumstance adverbial in Ll and L2 student academic presentations. *Journal of English for Academic Purposes, 8*(1), 55-68. https://doi.org/10.1016/j.jeap.2008.12.002

Zareva, A. (2016). Multi-word verbs in student academic presentations. *Journal of English for Academic Purposes, 23,* 83-98. https://doi.orgl10.1016/j.jeap.2016.07.001

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16921

Exercise parameters influencing exercise difficulty

Tanja Heck[a] and Detmar Meurers[b]
[a]Department of Linguistics, University of Tübingen, , tanja.heck@uni-tuebingen.de and [b]Department of Linguistics, University of Tübingen, , detmar.meurers@uni-tuebingen.de

How to cite: Heck, T.; Meurers, D. (2023). Exercise parameters influencing exercise difficulty. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16921

Abstract

Macro-adaptive systems aim to assign practice exercises to language learners that match their proficiency levels. While learner-dependent parameters of exercise difficulty need to be considered online, learner-independent parameters can inform an exercise's difficulty level in a resource-efficient offline procedure. We present an evaluation of learners' responses to form-based grammar exercises that aims to identify learner-independent exercise parameters affecting exercise complexity. The results indicate that the exercise type can yield coarse-grained complexity estimates, whereas exercise type specific features can inform more fine-grained estimates. For fine-grained estimates, we show that syntactic variants significantly impact exercise difficulty. Since there is strong variation between learning targets and learners with respect to the impact of different exercise parameters on a learner's performance, exercise difficulty can only be reliably determined if the exercises are created in a systematic way and by also considering characteristics of the learner.

Keywords: *ICALL, ILTS, exercise difficulty.*

1. Introduction

Learners' performance on language exercises is closely linked to exercise difficulty (Buckledee, 2008), which depends on learner-specific parameters on the one hand, and on exercise-specific parameters on the other hand (Pelánek et al., 2021). Macro-adpative systems that assign exercises to learners in a personalized manner for best possible learning outcomes, therefore need to consider both types of parameters when selecting an exercise (Liu et al., 2021). Learner-specific parameters, such as cognitive abilities or personal experience, are dynamic features and therefore need to be factored in online individually for each learner at the time of selecting an exercise (Kunichika et al., 2002). Exercise complexity, on the other hand, comprises learner-independent, static parameters of exercise difficulty, such as linguistic complexity of the textual material and characteristics of the exercise types, thus constituting a property of the exercise. In order to provide this meta-information to the exercise selection algorithm of a macro-adaptive system, exercise complexity can therefore be determined once offline before adding the exercise to the system's resources (Pandarova et al., 2019). Considering the stress that online calculations put on a system's performance, these offline calculations should take into account all learner-independent parameters of exercise difficulty in order to speed up the system's exercise selection process at runtime. It is thus necessary to not only identify parameters impacting exercise difficulty, but also to determine which of them are learner-independent.

Little is known about the impact of different exercise parameters on learner-independent exercise complexity and learner-dependent exercise difficulty. With our analysis of real-world learner data from two field studies, we shed light onto the relevance of a selection of exercise features of form-based, English grammar exercises in order to provide macro-adaptive systems with the means to more effectively and efficiently select exercises tailored to the individual learner.

2. Data

The evaluations are based on data collected from German 7th grade learners of English in the Interact4School (I4S) (Parrisius, Pieronczyk, et al., 2022; Parrisius, Wendebourg, et al., 2022) and the Digbindiff (Didi)[1] studies, which are based on the Intelligent Language Tutoring System FeedBook. The system offers exercises for practice of English as a second language, incorporating intelligent feedback provided as a learner works on the exercises. While both studies were conducted over the course of a school year, I4S focused on motivational aspects in a task based setting whereas Didi investigated the effects of user-adaptive exercise sequencing. For form-based grammar exercises, the FeedBook covers the seven exercise types Fill-in-the-Blanks (FiB), Single Choice (SC), Jumbled Sentences (JS), Categorization, Memory, Short Answers (SA), and Mark-the-Words (MtW). The exercises provide a total of 3,143 actionable elements, (N_{I4S}=1,140; N_{Didi}=2,003) such as blanks of FiB or SC exercises, chunks of JS exercises, elements to sort into a category, Memory pairs, answers to SA questions, or clickable words in MtW exercises. They are distributed across 11 distinct learning targets (N_{I4S}=9; N_{Did}i=4). While revising and re-submitting an exercise was possible in the studies, the evaluations only consider the first submission (N_{I4S}=153,596; N_{Didi}=120,431).

3. Evaluation

3.1. Exercise parameters impacting exercise difficulty

In order to determine those exercise parameters that are most predictive of exercise difficulty, we trained statistical models from the Python *scikit-learn* library to predict exercise difficulty based on a selection of exercise parameters. These parameters cover: a) generally applicable parameters including the number of actionable elements in the exercise or the length of an actionable element; b) exercise type specific parameters such as the number of distractors of SC exercises, the number of chunks of a JS exercise, the number of categories of a Categorization exercise, the number of pairs of a Memory exercise, or whether a FiB exercise requires the learner to determine the correct lemma in addition to transforming it into the correct form; and c) the exercise type itself. All predictors were encoded as numerical features. Difficulty of the actionable elements was operationalized as item difficulty scores obtained from an Item Response Theory model[2]. While the continuous scores served directly as outcome variables for the regression models, they were transformed into categorical values for the classification models. Since the FeedBook distinguishes three proficiency levels of learners, we applied the same amount of exercise difficulty levels. The thresholds between the three levels were determined through K-means clustering. All employed statistical models support feature ranking, which allows to easily determine their most predictive features. In order to identify the overall most predictive features, we added up the predictor ranks of all regression and those of all classification models, thus obtaining overall rankings for regression and classification, respectively.

[1] http://digbindiff.de

[2] The implementation is based on the Rasch model of the *TAM* package for R.

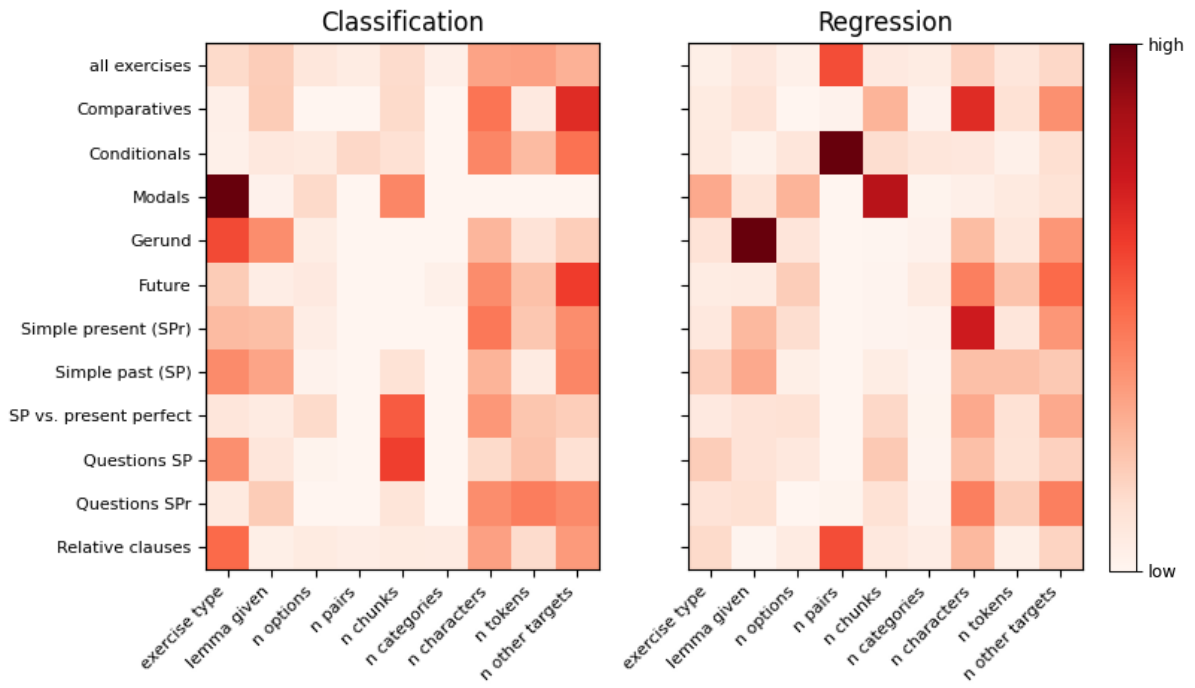


Figure 1. Feature importances in statistical models predicting exercise difficulty.

The heatmaps in Figure 1 assign colours of increasing darkness to parameters on the x-axis for learning targets on the y-axis the more important the parameter is for that learning target. They show that for the generally applicable parameters, the rankings are rather similar across learning targets for both the regression and the classification models. They all occupy ranks in the middle ranges, indicating that while they do not constitute the most informative of the evaluated parameters, their predictive power is rather constant and reliable across exercises. The exercise type specific features show considerably more variance across learning targets especially with regression, appearing at both extremes of the rankings. A general trend sees the exercise type as rather important for classification, whereas it ranks among the least predictive features for regression. Type-specific parameters hold more predictive power with those models. The regressions per exercise type, illustrated in the heatmap in Figure 2, highlight that the parameters applicable to only a particular exercise type indeed are more important for the respective exercises. This might indicate that the exercise type can inform coarse-grained difficulty estimations, while fine-grained distinctions require more detailed, type-specific parameters.

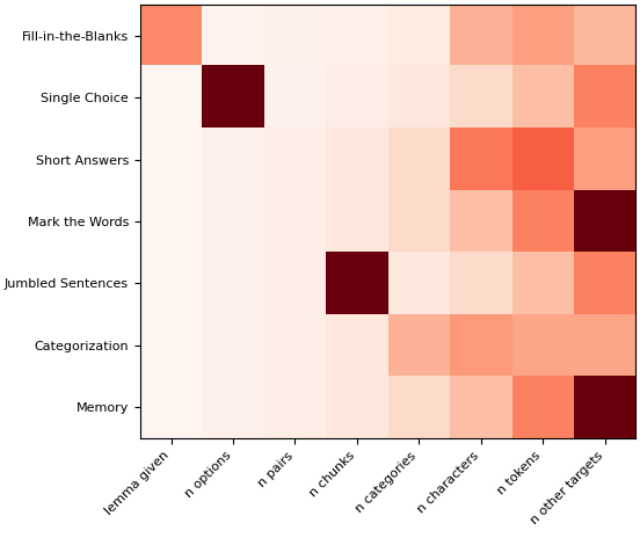


Figure 2. Feature importances for regression predicting exercise difficulty per exercise type.

3.2. Impact of syntactic variations on exercise difficulty

Since both classification and regression highlight the relevance of the generally applicable parameters, including linguistic complexity features such as token and character counts, we took a closer look at the impact of a range of linguistic complexity features on exercise difficulty. Didi's exercises of the learning targets *conditionals* and *relative clauses* were generated with the approach to systematic variability presented by Heck et al. (2022), so that they contain learner data for exercises with identical textual material and varying only in a selection of controlled, syntactic features. These variations target the *clause order*, *targeted clause* and *negation of clauses* for conditionals, and *clause order* for relative clauses. In order to determine their effect on exercise difficulty, we compared the distributions of difficulty scores across the different realizations of the variations. We tested for statistical significance with a two-tailed T-test, applying the commonly used threshold of $p<.05$ for statistical significance.

For the overall dataset, all effects were statistically significant, indicating that syntactic variations indeed impact exercise difficulty. In order to verify whether this is the case for all exercise types, we performed evaluations for the individual types. The violin plots given in Figure 3 illustrate that the results vary considerably across different exercise types, yet almost all effects are statistically significant. For the clause order of conditionals, exercises of almost all types are more difficult when putting the if-clause before the main clause. The effect, although not significant ($t=1.7796$, $p=.0760$), is inverted for Categorization exercises. With respect to the targeted clause, exercise items are slightly more difficult if the actionable element is in the if-clause rather than in the main clause with significant effects for all exercise types. Although we hypothesize that items simultaneously targeting both clauses are more difficult, the dataset does not contain according exercises. The question whether this variation of the exercise parameter makes a difference thus remains an open research question. Concerning negation, there is a statistically significant effect indicating that exercises are easiest if only the if-clause is negated, and most difficult when both clauses are negated. The only contradictory – and non-significant – effect appears with JS exercises between negation of the if-clause and of both clauses ($t=.1993$, $p=.8421$). For relative clauses, exercise items appear more difficult if the prompt gives the clause corresponding to the relative clause before that corresponding to the main clause. The effect is significant for all exercise types but Memory ($t=-.4771$, $p=.6333$).

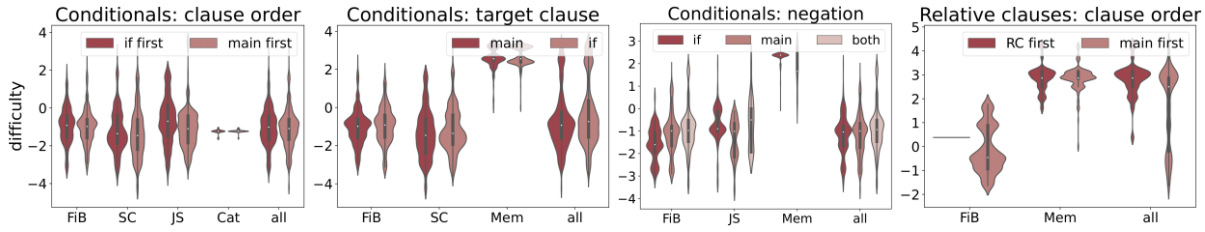


Figure 3. Difficulty distributions for exercises of syntactic variations.

3.3. Learner dependence of exercise parameter predictiveness

Since the exercise type holds predictive power for coarse-grained difficulty estimates, we investigated whether this is a global parameter for general exercise complexity or has to be determined on a per-learner basis. To this purpose, we determined the rankings of exercise types with respect to their difficulty for each learner and compared the distributions of ranking positions across learners. Following Pelánek et al.'s (2021) approach, we operationalized exercise difficulty as learners' performance on the exercises. More precisely, the rankings were created based on the ratio of incorrect to all submissions for an exercise item. If two exercise types obtained similar accuracies, they were both assigned the same rank.

The results are visualized in the heatmaps in Figure 4, where darker colours of a matrix cell indicate higher numbers of learners for which the exercise type is placed at the corresponding rank relative to the other exercise types. A single dark cell and white colour for all remaining cells of the row would indicate perfect agreement in

ranking for that exercise type among all learners; uniform colouring of a row would indicate highest possible diversity among learners. The heatmaps show that while there are differences between learners, there are definite tendencies as to what exercise types are most often solved incorrectly. The exact rankings are not identical for all learners, yet for most exercise types, the most frequent rank positions correspond to adjoining cells of the matrix, indicating that rough difficulty placements are similar for most learners. The results are clearerst for Fib and SC exercises, where the majority of learners make the least errors amongst all exercise types. MtW exercises constitute an interesting case as they feature two peaks at opposite ends of the ranking, indicating that they are among the least critical types for some learners, and among the most critical ones for others. JS exercises exhibit a similar trend, although they paint an overall more diverse picture with a number of learners also placing them in the middle ranking positions. SA exercises generally constitute the most difficult exercise type. In addition, the heatmaps differ considerably from one learning target to another, sometimes even reversing ranking positions. Assuming that learners perform better on exercises if they are more proficient in the skill that the exercise practices, this seems to indicate that the exercises of the dataset do not encode the same skill for an exercise type across all learning targets. It is therefore imperative to systematically create exercises so that they target the same skill for a particular exercise type, or else to also consider the skill when estimating an exercise's complexity.

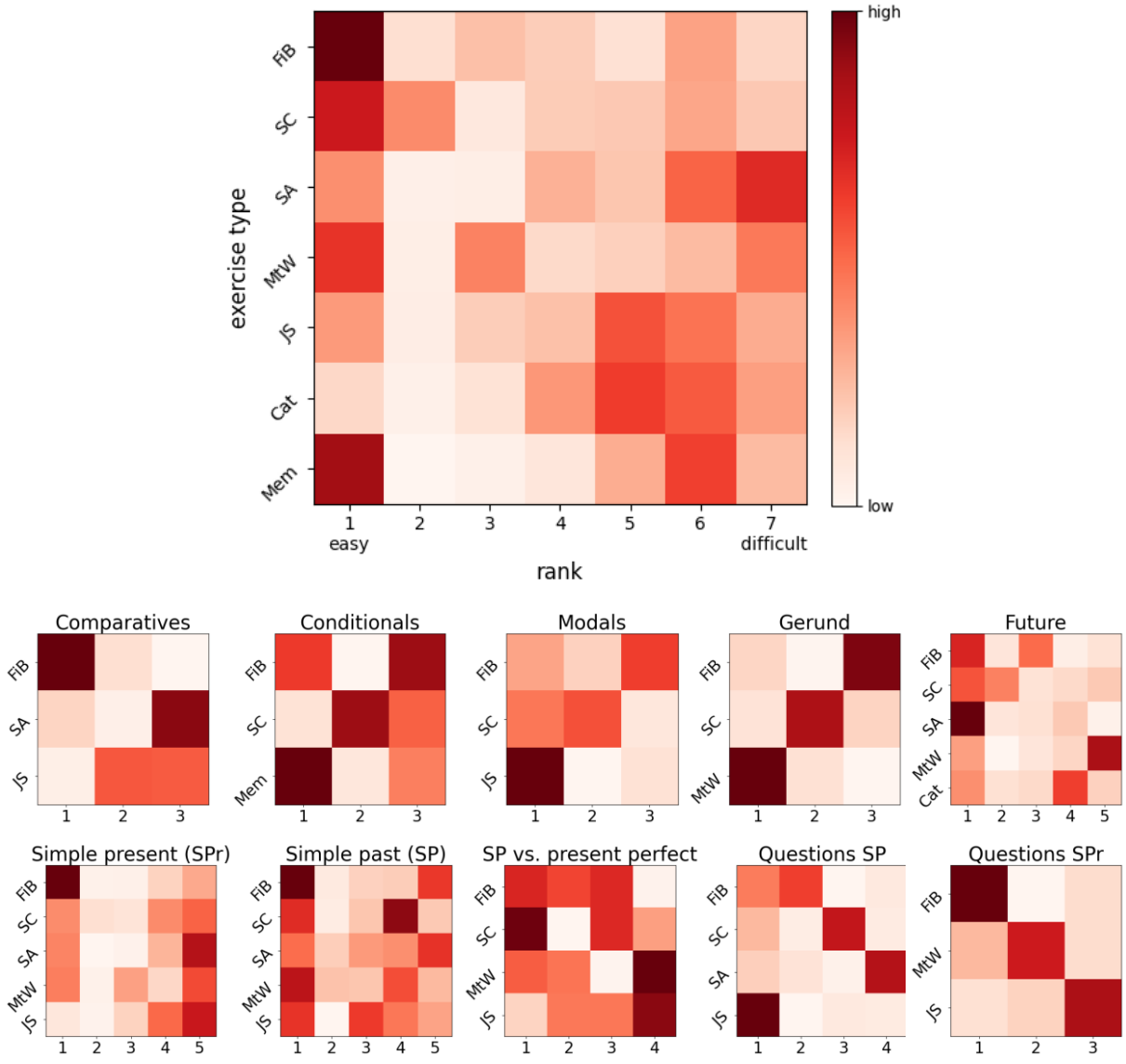


Figure 4. Distributions of learner-based exercise difficulty rankings

4. Conclusions

We presented an evaluation of exercise parameters with respect to their influence and learner-dependence on exercise difficulty. We found that the exercise type is indicative of coarse-grained difficulty estimates, while exercise-type specific parameters can yield more fine-grained predictions. Although the parameters hold some general predictive power, even coarse estimates are best based on exercise features in conjunction with learner characteristics. Syntactic variants do have an impact on exercise difficulty, so that macro-adaptive systems should take these linguistic features into account when calibrating exercise difficulty. While the approach presented by Pandarova et al. (2019) could be extended to consider parameters of syntactic variations, also taking learner characteristics into account requires maintaining and consulting a learner model at runtime.

References

Buckledee, S. (2008). Motivation and Second Language Acquisition. In Z. Dörnyei & R. W. Schmidt (Eds.), *ELOPE: English Language Overseas Perspectives and Enquiries* (Vol. 5, Issues 1–2, pp. 159–170). Second Language Teaching & Curriculum Center, University of Hawai'i at Mānoa. https://doi.org/10.4312/elope.5.1-2.159-170

Heck, T. & Meurers, D. & Nuxoll, F. (2022). Automatic exercise generation to support macro-adaptivity in intelligent language tutoring systems. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, Research-publishing.net, pp. 162-167. https://doi.org/10.14705/rpnet.2022.61.1452

Kunichika, H., Urushima, M., Hirashima, T., & Takeuchi, A. (2002). *A Computational Method of Complexity of Questions on Contents of English Sentences and its Evaluation*. 97–101. https://doi.org/10.1109/CIE.2002.1185873

Liu, Q., Shen, S., Huang, Z., Chen, E., & Zheng, Y. (2021). A survey of knowledge tracing. *ArXiv Preprint ArXiv:2105.15106*.

Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R. D., & Brefeld, U. (2019). Predicting the Difficulty of Exercise Items for Dynamic Difficulty Adaptation in Adaptive Language Tutoring. *International Journal of Artificial Intelligence in Education*, *29*(3), 342–367. https://doi.org/10.1007/s40593-019-00180-4

Parrisius, C., Pieronczyk, I., Blume, C., Wendebourg, K., Pili-Moss, D., Assmann, M., Beilharz, S., Bodnar, S., Colling, L., Holz, H., & others. (2022). *Using an Intelligent Tutoring System within a Task-Based Learning Approach in English as a Foreign Language Classes to Foster Motivation and Learning Outcome (Interact4School): Pre-registration of the Study Design*. PsychArchives. https://doi.org/10.23668/psycharchives.5366

Parrisius, C., Wendebourg, K., Rieger, S., Loll, I., Pili-Moss, D., Colling, L., Blume, C., Pieronczyk, I., Holz, H., Bodnar, S., & others. (2022). *Effective Features of Feedback in an Intelligent Tutoring System-A Randomized Controlled Field Trial (Pre-Registration)*. PsychArchives. https://doi.org/10.23668/psycharchives.8152

Pelánek, R., Effenberger, T., & Čechák, J. (2021). Complexity and Difficulty of Items in Learning Systems. *International Journal of Artificial Intelligence in Education*, *32*, 1–37. https://doi.org/10.1007/s40593-021-00252-4

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16981

FOCUS - For a functional, digital, and critical literacy

Beatrice Johansson[a] and Qarin Franker[b]

[a]COO at Dispurse Foundation, beatrice@dispurse.org and [b]Senior Adviser at Dispurse Foundation, qarin.franker@gmail.com

How to cite: Johansson, B.; Franker, Q. (2023). FOCUS for a functional, critical and digital literacy. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16981

Abstract

This article describes the development of a new educational application, FOCUS, for women who are learning to read and write for the first time in their lives. Their driving force is often to manage their current everyday life better but also to change their future living conditions. Initially the Dispurse Literacy Program and the applications´s prerequisites and previous development are described. Examples of how the content and design has been changed with respect to both technical and pedagogical requirements are also given. The theoretical framework, the Resource Model, is then presented followed by a section where four pedagogical questions guiding the development of the app are examined. Finally, a short section follows where challenges and future possibilities for digitally based literacy education are highlighted.

Keywords: *adult literacy program, educational application, emergent digital literacy, Resource Model.*

1. Dispurse Literacy Program

There are about 763 million people in the world that lack basic literacy, two thirds of these are women (UNESCO, 2023a). For women with no or limited previous experience of formal education, taking an active part in their community can be challenging. The educational opportunities are often few and due to geographical, economical, and social factors, attendence, as well as access to digital tools, can be difficult.

The Swedish organization, Dispurse, has been offering a free and digitally based literacy program in remote rural areas of Peru since 2017. The program has been adapted for Quechua speaking women who have not had the opportunity to go to school. A majority of the learners are bilingual and also speak Spanish. To date, more than 2,000 women have participated in the Dispurse Literacy Program (DLP). To reach as many learners as possible Dispurse cooperates with stakeholders, such as mayors in local communities, regional governments, and the Ministry of Education in Peru at a national level.

A central part of DLP consists of the digital application FOCUS, which is developed to be used on tablets and works without access to wifi. It aims to offer a basic literacy education that is functional, digital, and critical (Dispurse, 2023). The app is combined with complementary learning materials that have been developed to supplement and deepen the learners digital work. These materials should be considered as an integrated whole as the learners can work both individually and together in groups, with different literacy practices in combination with their work in FOCUS. This group work is led by trained facilitators who visits the learners in their communities on a regular basis.

Each learner´s progress in DLP is also visualized for facilitators and administrators in the websystem, DispurseAdmin. The progress in FOCUS is synchronized by the facilitator using their mobile phones in places

with access to wifi. In places without access to wifi facilitators use our application DispurseSynch which enables local synchronization of the learners progress.

2. From a technology driven to a user-driven development

The development of FOCUS began in 2011 in Sweden. During the years 2014-2017 the application was tested by target groups in four Latin American countries. From 2017 and onwards the development has focused on learners in Peru where a local branch of Dispurse also was established.

Initially, the content of FOCUS consisted of exercises that enabled learners to work with codebreaking practises, focusing on what the Council of Europe (2023) describes as *technical literacy*, with exercises mainly increasing the learners phonological awareness by reading and writing letters and short, wellknown words. But also exercises that aimed to increase the learners vocabulary using *individual cognitive processes* of language learning. A focus on codebreaking practises was also found by Kim & Kwon (2012) when they reviewed available mobile apps for language learning.

Based on the testing of FOCUS during 2014-2017, we decided to contextualize the content more as well as integrate a content that enabled the learners to work with both *communicative and social processes* and *individual cognitive processes* of language learning. This highlights an important change for Dispurse Literacy Program (DLP). It changed the view of literacy as a mainly autonomous skill to a view of literacy as an active process situated in social practises, interdependent on different cultures and contexts (Street, 1993). This broader view of literacy also corresponds to UNESCO's definition of literacy (2023b) as "the ability to identify, understand, interpret, create, communicate, and compute, using printed and written materials associated with varying contexts".

An early example from FOCUS of the shift from an autonomous to a more contextualized view of literacy is vizualised in both content and user interface in the older and newer versions of FOCUS (Figures 1 and 2).

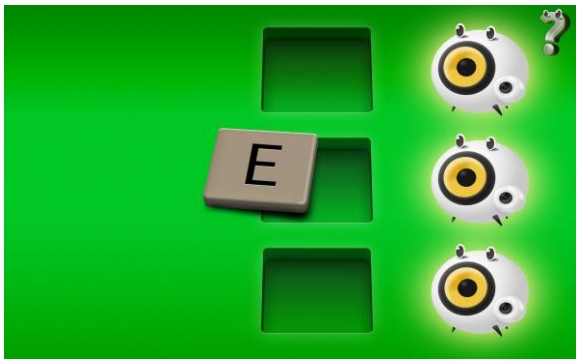


Figure 1. Linking graphem and phonem in older versions without context.

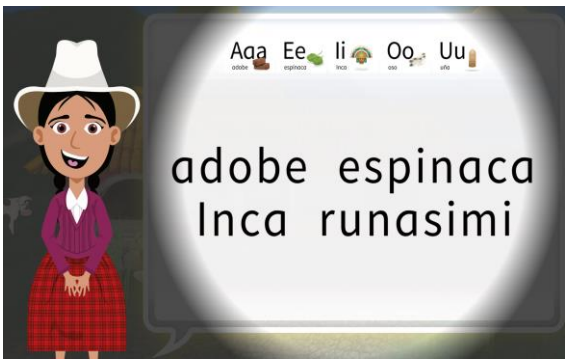


Figure 2. Linking graphem and phonem in newer versions withfamiliar words in context.

Learner feedback also led to an integration of everyday concepts and language into the app, for example the use of ID-cards, an electricity bill, and recipes for nutritious food. An intensive development of complementary teaching and learning materials to further strengthen the functional literacy perspective started in parallel.

In 2022, Dispurse initated an external research-based evaluation. The study assesses the potential usefulness and meaningfulness of the literacy program, as well as its correlation with learning outcomes. The evaluation process is theoretically supported by Franker (2016) and Luke & Freebody (1999) and relies on specifically constructed instruments inspired by PIAAC in OECD (2019). Their preliminary results will be presented in 2024.

3. A Resource Model for basic literacy programs

The Resource Model (Figure 3) serves as a theoretical and didactic starting point for the further development of FOCUS and the complementary learning materials. The model was initially created by Freebody and Luke' and called ‘The Four Resources Model’ (1999). Our version is supplemented with the learner resources, adapted to initial literacy education for adults, and also visualized by Franker (2016) as a puzzle with the learners in the center. In the change of the model, Barton´s (2007) view on literacy as a set of social practices, a symbolic system used for communicating and for representing the world to ourselves, and Janks (2010) critical perspectives on literacy and power, was integrated.

The model shows how the linguistic, literacy-related, and socio-cultural resources that each individual has already acquired become the starting point for their practical literacy work and influences their development within the four different literacy practices; code-breaking practice, meaning making practice, text analyzing practice, and the text using practice.

By solving real problems while working with the four different literacy practices, the learners crack, understand, use, and question the script code. By analyzing interesting words and texts together, using everyday texts, and copying and creating new texts together their knowledge of letters, words, and context increases. When these varied focuses are integrated into their daily work, the learners develop a useful literacy foundation which can also provide a solid foundation for further studies (Franker, 2016).

Figure 3. The Resource Model – Learner resources and literacy practices interact and drive the learning processes (Franker, 2016).

4. Four questions for developing the literacy application

Having the Resource Model and its focus on learners resources, four overarching questions guide the development of FOCUS, each discussed and illustrated with examples from the app.

How do we show respect for the learners linguistic and cultural experiences?

A challenging question is how the design and content can show respect for the learners' linguistic and cultural experiences (Figure 3). Done with respect and in dialogue with the learners possible stereotyping could be avoided. Through the use of Maria (Figure 4), as well as the design of her environment, referencing a well-known context, in which the learners can recognize themselves, is enabled. The road signs makes use of Quechua sounding names and Maria also wears the 'white hat', a cultural marker of the Quechua population in northwestern Peru.

Figure 4. Maria welcomes the learners to the FOCUS app.

To take advantage of the learners linguistic resources as well as offer the learners support in their first language, a prerequisite for our facilitators is that they can speak the languages spoken in the area where the program is implemented (Winlund, 2020). References to the Quechua language are used as often as possible in the choice of content in FOCUS. Throughout the learning process, Maria asks questions to enhance group discussions whilst highlighting their lingustic resources. She encourages them to use the language that feels most comfortable for them.

How do we create recognition and interactivity?

Recognition is a central factor for learning and in order to enhance the learners possibility to recognize and comprehend the materials we strive to use authentic and often local, well-known materials.

The first of our four themes in DLP is about Identity and the text we use is from the personal ID card, a familiar and essential text artefact in Peru and important for the learners in their daily lives. Another theme builds on our learners oral proficiency and listening comprehension in Spanish as well as their traditions of storytelling. This highlights the importance of oral practises not only written ones, which are often seen as the most prestigous ones in educational settings (Janks, 2010).

Figure 5. The personal ID-card in FOCUS.

Figure 6. Group work with the personal ID-card.

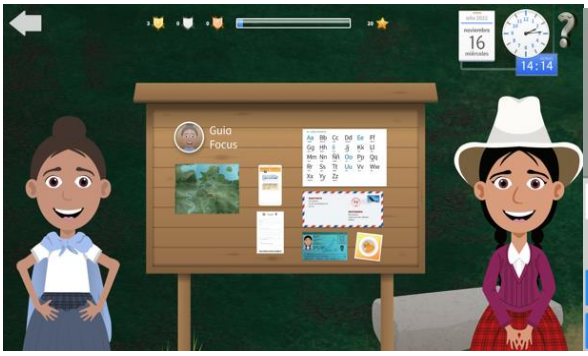
We work to provide the learners with content they can work with in groups, connecting FOCUS to group work sessions.

Interactivity is a criteria linked to user experience in the framework for evaluating language learning apps found by Rosell-Aguilar (2017). The overarching structure of each module in FOCUS contains a collaborative part with a primary function of promoting interactivity between learners, as well as giving learners access to scaffolding from the facilitator and other group members (Gibbons, 2014).

How can we make the learners active and co-creative?

The learners work in FOCUS starts with making their own avatar together with Maria. It seems that this initial act, when they can choose their own clothes, type of hat or no hat, and colours of hair and eyes, make the learners feel comfortable and ready to start their work.

There is also a personal module where each individual learner has their own billboard. Here, they get an overview of their own progress and their own path in the program. They can practice writing or reading important names, words, messages, and texts that relate to them. They can also at an early stage use the keyboard with speech synthesis to write their own short texts, listen to them, and pay attention to reading and writing directions. Different text types introduced in FOCUS can also be found here and the learner can practise navigating in a personal ID-card or write a recipe as well as writing short text-messages and using emojis.



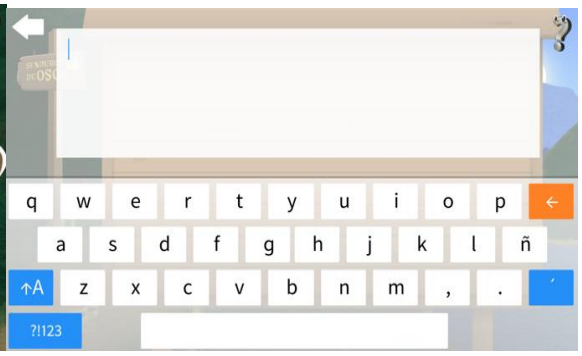


Figure 7. The personal module.

Figure 8. The keyboard at the personal module.

How can we challenge prevailing norms in society with the help of design?

Norms and conventional ways exist on how to use written language. This use is regulated by social institutions, and defined by culture which creates norms in society concerning what languages and literacies are the most prestigious and which are not. For us, the main dilemmas when developing the DLP is still the choice of language and meaningful content.

The story of Maria and her world highlights an indigenous woman living in a rural area which not is the norm in any common learning materials. The design and development was based on a thorough dialogue and collaboration with learners, facilitators, and local stakeholders with the aim to strengthen our learners identities and local culture.

We found using Quechua in the app difficult due to the many local variations. Choosing Spanish to give our learners access to the dominant language and at the same time promote Quechua was, and still is, what Janks (2010) calls an access paradox. By using Spanish its dominance is reproduced, but by denying the learners access to it their marginalization will only continue.

We strive to offer a content that is perceived as meaningful and useful for the learners in their everyday lives and that can enhance their participation in society. We have today choosen to include a common procedure and instrument of participation in Peru which is to understand the design and how to fill in a formal request to meet

the local mayor. Another useful literate strategy included in the content is to know how to contact an organization for help when living in a violent relationship.

5. Challenges and future possibilities

We are confident that computer technologies can play an important role in providing literacy education, including digital skills and language learning for all who need it. Technology needs to be grounded in the local contexts and needs of the learners. It should be designed with respect to the learners languages, cultures, and literacy related resources.

For this to happen, we see the need for further research into how CALL-technologies can be designed and adapted for adult learners with no or very limited previous experience of formal education. Research studies could also have a stronger focus on how technology can support innovative and even more critical pedagogies. We also look forward to see more collaborations and sharing of best practises in how CALL technologies can support the low literate target group which has been the focus of this article.

We hope that our experiences can serve as inspiration for others working to provide basic adult literacy. Lack of education hampers individual and social development and increases inequality between people. As Freire (1970) pointed out, people that are the poorest, most oppressed, and least powerful also tend to be the ones that lack literacy education. The need for innovative CALL entrepreneurs and researchers in this field is still immense.

References

Barton, D. (2007). *Literacy. An introduction to the ecology of written language*. (2nd ed.). Blackwell Publishing.

Council of Europe. (2023). *Literacy and second language learning for the linguistic integration of adult migrants* (Laslliam). Strasbourg: Council of Europe Publishing.

Dispurse. (2023). *FOCUS- guide*; https://dispurse.org/media/zyulsxh4/focus_guide_18-final_eng.pdf

Franker, Q. (2016). *Grundläggande litteracitet för nyanlända ungdomar.*

Freire, P. (1970). Pedagogy of the Opressed. (Pedagogia do Oprimido, 1968) ISBN 978-0-8264-1276-8

Gibbons. P. (2014). *Scaffolding language, Scaffolding learning.* Teaching English Language Learners In The Mainstream Classroom. Heinemann Educational Books.

Janks, H. (2010). *Literacy and power*. New York: Routledge

Kim, H. & Kwon, Y. (2012). Exploring smartphone applications for effective Mobile- Assisted Language Learning. *Multimedia-Assisted Language Learning*, 16(1), p. 31-57.

Luke, A., & Freebody, P. (1999). *A map of possible practices: further notes on the four resources model*. Practically Primary; v.4 n.2 p.5-8; June 1999, 4(2), 5–8.

OECD. (2019). Skills Studies. The Survey of Adult Skills. https://www.oecd.org/skills/piaac/publications/Survey_of_Adult_Skills_Reader's_Companion_3rd_Edition_ENG.pdf

Rosell–Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *Calico Journal* vol 34.2 p. 243-258. https://doi.org/10.1558/cj.27623

Street, B. V. (1993). Introduction: The New Literacy Studies. In B. V. Street (Ed.), *Crosscultural approaches to literacy* (pp. 1-21). New York: Cambridge University Press

UNESCO. (2023a). *What you need to know about literacy*. https://www.unesco.org/en/literacy/need-know

UNESCO (2023b). *Literacy. Definition* https://uis.unesco.org/node/3079547

Winlund, A. (2020, February 5). Emergent literacy instruction: 'continua of biliteracy' among newly immigrated adolescents. *Language and Education*, *34*(3), 249–266. https://doi.org/10.1080/09500782.2019.1701006

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16956

Investigation of PoodLL ReadAloud in Moodle to reduce the foreign language learning anxiety in English and Japanese language classes

Maki Terauchi Ho[a]

[a]Department of Human Science, Obihiro University of Agriculture and Veterinary Medicine, terauchi@obihiro.ac.jp

How to cite: Ho, M. T. (2023). Investigation of PoodLL ReadAloud in Moodle to reduce the foreign language learning anxiety in English and Japanese language classes. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16956

Abstract

The advantages of learning a foreign language are widely acknowledged, and it is commonly understood that practice is essential for acquiring proficiency. However, learning another language has various obstacles, such as language learning anxiety which hinders students' learning capabilities. There are many studies to attempt to identify the anxiety provoking factors (Cheng et al., 1999; Khattak et al., 2011; Onwuegbuzie et al., 2000; Sellers, 2000; Zheng, 2008). Recent technological advancement offers many potential methods to empower student learning without anxiety. This study aimed to investigate one such tool, PoodLL ReadAloud activity in Moodle to enhance speaking practice. This study compared the anxiety levels experienced while speaking in foreign language classrooms with those experienced when using PoodLL on a computer, employing questionnaires (Short Foreign Language Classroom Anxiety Scale and PoodLL Anxiety Scale). The anxiety level associated with using PoodLL was lower in first-year and second-year university classrooms, but not among graduate students. Additionally, feedback from the students using PoodLL revealed a positive practice experience, technical difficulties, and scepticism of AI grading. In conclusion, PoodLL shows promise as a tool for practicing speaking skills, however, further research is needed to assess its impact on skill improvement.

Keywords: *PoodLL, Moodle, speaking task, foreign language learning anxiety.*

1. Introduction

Speaking a foreign language offers numerous advantages, but individuals often encounter difficulties when they embark on learning another language. For example, the fear of making errors and feeling embarrassed, especially in speaking, can impede concentration and hinder optimal performance. Researchers have identified this anxiety as a common occurrence in foreign language classrooms and have conducted numerous studies to explore the contributing factors (Cheng et al., 1999; Khattak et al., 2011; Onwuegbuzie et al., 2000; Sellers, 2000; Zheng, 2008). Despite various studies, there is a scarcity of research focusing on types of less anxiety-provoking speaking practices to enable learners to reach their full potential. The technological advancements, including Artificial Intelligence (AI), offer significant potential for enhancing speaking practice. This study aimed to assess whether using PoodLL ReadAloud activities in speaking practice would be less anxiety-provoking than in regular classrooms, and gather students' feedback on their experience with PoodLL.

There are several studies to investigate and indicate the anxiety factors in foreign language classrooms (Ohata, 2005, Khattak et al., 2011, Effiong, 2015). Students reported that they constantly felt they were evaluated by teachers and peers, sometimes even by themselves (Ohata, 2005), and the fear of making mistakes in class could raise their anxiety level (Khattak et al., 2011). Even the teachers' dress code, their age, tone of voice and gender would provoke anxiety in the classroom (Effiong, 2015). Anything perceived as unfriendliness or strictness by the students could raise their anxiety level, and the students experienced stress, anxiety or nervousness during the foreign language class.

A key factor to influence the anxiety level in the language classroom is the students' level of confidence. Cheng et al. (1999) claimed the learners' confidence was a good predictor of their anxiety level. If the learners are confident in their speaking and writing abilities in the target language, they score lower in Foreign Language Classroom Anxiety Scale (FLCAS). MacIntyre et al. (1997) found that low-confidence learners tend to have negative expectations of language learning itself (MacIntyre et al., 1997). Increased anxiety among learners negatively impacts their willingness to engage in classroom communication, consequently impeding their learning opportunities (Cheng et al., 1999). Therefore, the students who are confident in the target language are less anxious and more likely to participate in classroom activities.

Some researchers use audio or video recording to alleviate speaking anxiety. Young & West (2018) analyzed the studies using Asynchronous Multimedia-based Oral Communication (AMOC), such as video conferences, vlogs, YouTube videos, turn-based video conversation, and computer-mediated communication. Compared to the synchronous communication activities, AMOC showed promise in helping fluency, accuracy, and pronunciation.

This study investigates speaking practice using PoodLL ReadAloud activity among university students. There are 3 questions to be considered in this study.

1. Do the students have lower anxiety levels using PoodLL than in the classroom?
2. How many assignments did they complete, and what is the reason for the completion/incompletion?
3. What was the overall impression of PoodLL? Are there any difficulties?

2. Method

2.1. PoodLL ReadAloud

PoodLL, a Moodle plugin, enriches the platform with audio and video recording tools and activities to enhance reading and pronunciation skills. In this study, students used the 'ReadAloud' activity to practice speaking. Each week they could choose to shadow read with headsets (simultaneous listening and speaking) or read aloud. They also had the option to listen to a model speech before recording.

2.2. Participants and context

This study included 8 non-Japanese graduate students in Beginning Japanese, 19 graduate students (including 3 non-Japanese students) in English for Science, 26 second-year students in English III, and 26 first-year students in English I. Japanese class participants received Japanese assignments, while those in English classes received English assignments. They completed 10 assignments before responding to the questionnaire.

The passages, selected based on class relevance and language proficiency, ranged from 150 to 300 words. Students had 1 minute 30 seconds to 3 minutes for reading. AI grading assessed their reading speed and pronunciation accuracy, providing instant feedback after recording.

2.3. Questionnaires (Short Foreign Language Classroom Anxiety Scale and Poodle Anxiety Scale)

I opted for questionnaires as the data collection method due to their minimal impact on students' schedule and their efficiency in gathering responses from many participants. The first questionnaire took place seven to eight weeks after the term started to prevent confusion with the beginning of the academic year or college life for first-year students. The second questionnaire followed ten weeks of ReadAloud assignments. Students were informed

of the study's purpose and their option to abstain without academic penalty. Those who agreed to these conditions voluntarily joined the study.

First, participants completed the Short Foreign Language Classroom Anxiety Scale (SFLCAS) by Botes et al. (2021) with its Japanese version (Yashima et al., 2009). To reduce respondent stress and align with the study's focus on comparing anxiety related to the PoodLL and classroom settings, the SFLCAS with its concise eight questions was used instead of the lengthier FLCAS (Horwitz et al., 1986), which comprises 33 questions. SFLCAS was adapted by modifying 'classroom' to 'ReadAloud' to create the Poodle Anxiety Scale (PAS). Responses were rated on a 1-5 Likert scale (higher scores indicating higher anxiety). Additionally, PAS included three open-ended questions to probe assignment completion reasons, evaluate the PoodLL ReadAloud activity, and gather unanticipated insights regarding PoodLL. These questions were added to uncover any positive or negative reactions to the PoodLL activity, rather than solely focusing on anxiety-related psychological impacts.

3. Results

The study compared the means of SFLCAS and PAS for each classroom, as detailed in Table 1. Notably, the means of PAS were consistently lower than those of SFLCAS. However, the means among graduate students did not exhibit statistical significance. Additionally, due to an insufficiently small sample size, responses from the Japanese language class were excluded from statistical analysis.

Table 1. The comparison of the means of SFLCA and PAS.

Student Groups	SFLCAS	PAS	p-value
1st year (N=26)	28.43 (SD=5.91)	25.03 (SD=6.40)	P=0.048
2nd year (N=26)	29.23 (SD=5.29)	23.65 (SD=7.48)	P=0.005
Graduate (N=19)	28.00 (SD=7.26)	27.42 (SD=8.08)	P=0.817

On average, the students completed 8.44 out of 10 assignments. Regarding the reason for their completion/incompletion of the assignment, the individuals who did only a few or none of the assignments commented that they simply forgot to do it. No students mentioned having a negative reaction to the activity. Among those who completed the most assignments, approximately half commented that they felt they were obligated to do so because it was an assignment. However, the rest of the students responded as they thought it would improve their speaking skills and viewed them as valuable opportunities for practice. A few students found the activities enjoyable, either due to their interest in the material or by likening the assignment to a game.

Additional feedback regarding the ReadAloud activity included both positive remarks and technical problems, as well as apprehensions regarding AI grading. Many students, including non-Japanese, found the speaking practice with PoodLL was highly beneficial for their language study. They noted improvements in vocabulary, intonation, and sentence structure, ultimately enhancing their speaking confidence. They regarded it as a valuable opportunity for practicing spoken language skills. For example, one second-year student commented "I can practice speaking English by it [PoodLL]. It is a good opportunity to speak English for me." However, some students encountered technical difficulties. They reported inadequate time to complete reading the assigned passages. Other students faced issues with their microphones, preventing them from recording their speech and thus, impeding their assignment completion. Furthermore, some students expressed their concern about the AI grading system. They believed that they consistently received low grades, regardless of their efforts, leading to uncertainty about the accuracy of AI grading abilities.

4. Discussion

In this study, the implementation of PoodLL ReadAloud activity presented some advantages for both educators and students. The teacher found it straightforward to assign appropriate tasks and grade automatically. Additionally, students benefit from a user-friendly platform that facilitates speaking practice and provides prompt feedback. The ease of use not only saved educators valuable time but also empowered students to engage in language learning, resulting in positive feedback for gaining confidence through practice.

The use of PoodLL reduced anxiety levels in undergraduate classrooms compared to traditional speaking tasks. Some students tend to disengage from anxiety-inducing tasks, leading to limited practice opportunities that can hinder their learning (Cheng et al., 1999). PoodLL provides a low-anxiety environment for speaking practice, making it more appealing to students. Nevertheless, it's important to note the absence of a significant difference in graduate student anxiety levels. Further investigation and analysis are needed to better comprehend anxiety reduction dynamics across various educational contexts.

While the AI grading system in PoodLL offers convenience for instructors dealing with a large number of students, it warrants refinement. Notably, beginner to intermediate students often received excessively low scores for pronunciation accuracy, suggesting that word recognition may be overly sensitive. Such low scores could potentially discourage students from engaging in conversations with others (Zheng, 2008). To address this, manual score adjustments were made throughout the term, with explanations to students about potential inaccuracies in machine grading. Nonetheless, the provision of instant feedback remained a valuable motivational tool for learners (Swanson & Schlig, 2010). Therefore, system improvement could benefit from the addition of a feature allowing for the adjustment of grading sensitivity.

Consideration of technology accessibility is vital in class design (Young & West, 2018). PoodLL, a paid Moodle plugin, may not be accessible to all institutions. However, for students, usability is relatively simple, requiring only internet access and a microphone, as long as the institution can offer the module. Therefore, it is worth considering its incorporation into Moodle if it is accessible.

5. Conclusions

This study investigated the effectiveness of PoodLL ReadAloud activity in Moodle for reducing anxiety during speaking activities. While technical challenges and concerns about AI grading were noted, the results of the questionnaires demonstrated a reduction in anxiety levels. However, the study did not measure the enhancement of their confidence levels or assess the actual impact on students' speaking skills. A related study by Sun (2012) examined speaking practices using voice blogs but found no significant improvements in pronunciation, language complexity, fluency, or accuracy. Therefore, future research should incorporate an investigation of confidence levels and assessments of speaking skills. Furthermore, the absence of significant differences in anxiety levels among graduate students requires further investigation, indicating the complexity of anxiety in language learning. Moreover, a larger sample size is required for conducting a more comprehensive investigation among Japanese language learners. Consequently, the PoodLL holds promise for facilitating speaking practice in foreign language classrooms, but a more comprehensive examination of its impact on speaking fluency, pronunciation, accuracy, and language complexity is needed in future studies.

References

Botes, E., van der Westhuizen, L., Dewaele, J.-M., MacIntyre, P., & Greiff, S. (2022). Validating the short-form foreign language classroom anxiety scale. *Applied Linguistics*, *43*(5), 1006-1033. https://doi.org/10.31234/osf.io/x8mcg

Cheng, Y., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*(3), 417-446. https://doi.org/10.1111/0023-8333.00095

Effiong, O. (2015). Getting them speaking: Classroom social factors and foreign language anxiety. *TESOL Journal*, *7*(1), 132-161. https://doi.org/10.1002/tesj.194

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125-132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Khattak, Z. I., Jamshed, T., Ahmad, A., & Baig, M. N. (2011). An investigation into the causes of English language learning anxiety in students at AWKUM. *Procedia - Social and Behavioral Sciences*, *15*, 1600-1604. https://doi.org/10.1016/j.sbspro.2011.03.337

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, *47*(2), 265-287. https://doi.org/10.1111/0023-8333.81997008

Ohata, K. (2005). Potential sources of anxiety for Japanese learners of English: Preliminary case interviews with five Japanese college students in the U.S. *Teaching English as a Second or Foreign Language*, *9*(3), 1-21.

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research*, *94*(1), 3-15. https://doi.org/10.1080/00220670009598738

Sellers, V. D. (2000). Anxiety and reading comprehension in Spanish as a foreign language. *Foreign Language Annals*, *33*(5), 512-520. https://doi.org/10.1111/j.1944-9720.2000.tb01995.x

Swanson, P. B. & Schlig, C. (2010). Improving second language speaking proficiency via interactional feedback. *International Journal of Adult Vocational Education and Technology*, *1*(4), 17-30. https://doi.org/10.4018/JAVET.2010100102

Sun, Y. C. (2012). Examining the effectiveness of extensive speaking practice via voice blogs in a foreign language learning context. *CALICO Journal*, *29*(3), 494-506. https://www.jstor.org/stable/calicojournal.29.3.494

Yashima, T., Noels, K., Shizuka, T., Takeuchi, O., Yamane, S., & Yoshizawa, K. (2009). The interplay of classroom anxiety, intrinsic motivation, and gender in the Japanese EFL context. *Journal of Foreign Language Education and Research*, *17*, 41-64.

Young, E. H., & West, R. E. (2018). Speaking practice outside the classroom: A literature review of asynchronous multimedia-based oral communication in language learning. *The EUROCALL Review*, *26*(1), 59. https://doi.org/10.4995/eurocall.2018.8599

Zheng, Y. (2008). Anxiety and second/foreign language learning revisited. *Canadian Journal for New Scholars in Education/Revue Canadienne Des Jeunes Chercheures et Chercheurs En Education*, *1*(1). 1-12.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16862

Machine translation use in the English as a Foreign Language (EFL) classroom

Anthony Young[a]
[a]Literature Department, Aichi University, , anthony@vega.aichi-u.ac.jp

How to cite: Young, A. (2023). Machine translation use in the English as a Foreign Language (EFL) classroom. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16862

Abstract

Machine Translation (MT) increasingly has the potential to facilitate or impede second language learning in the English as a Foreign Language (EFL) classroom. To understand teacher and EFL learner perspectives about its use, two surveys of 20 university professors (10 non-native English teachers and 10 native English teachers) and 139 EFL students were carried out. To measure MT's capacity to promote language awareness, a comparative study with an online dictionary was done, using two translation tasks, a posttest, and a post-questionnaire. The preliminary survey results revealed variations in how learners and teachers perceive text-based machine translation as a valuable tool for language learning and how much they felt its use by students in educational settings should be accepted. Analysis of the posttest also found significant differences in the capacity of DeepL (MT network) to promote language awareness compared to the online dictionary, which aligned with the students' post-questionnaire feedback. These results emphasise the need for additional research and workplace dialogues going forward, regarding the incorporation and supervision of MT in L2 instruction.

Keywords: *machine translation, online dictionary, English as a Foreign Language, task-based learning.*

1. Introduction

Continued advancements in Machine Translation (MT) have likely escalated its impact on second language (L2) education. Rule-based translation systems have given way to neural MT networks like Google Translate and DeepL, which train themselves to produce increasingly accurate output (Jolly & Luiane, 2022). Such translation tools are readily accessible to L2 learners online and can be used for various academic and personal purposes. Previous studies have shown that MT helps reduce the cognitive burden of translations (Baraniello et al., 2016) and encourages self-directed learning (Godwin-Jones, 2015; Wong & Lee, 2016). At the same time, there are still ongoing debates among educators as to whether MT use by learners should constitute cheating (Jolly & Luiane, 2022), or if it turns learners into mere passive recipients of information (Innes, 2019). These are important concerns that need to be addressed. The aim of this study was to compare teacher and EFL learner perspectives on MT use, and to measure the capacity of MT to promote learning in a task-based environment, compared to an online dictionary. The research questions were as follows:

1. To what extent are the perspectives of teachers and EFL students similar or different, regarding MT use for learning?

2. To what extent are text translations produced by EFL students using MT or an online dictionary similar or different?

3. Is language awareness during a translation task promoted more by MT use or by using an online dictionary?

2. Method

For this study, data was collected over the period of one semester. An initial survey (Appendix A) regarding MT use for L2 learning was done first, involving 20 university professors (10 non-native English teachers and 10 native English teachers) and 139 EFL university students. A comparative case study of the MT network DeepL and an online dictionary jisho.org was then carried out to measure the capacity of both online tools to promote language awareness in a task-based learning environment. This study involved 22 EFL students from an academic writing class carrying out two translation tasks, a posttest, and a post-survey (Appendix B). To answer research question 1, the results of the initial survey (including Likert-scale responses and follow-up explanations) for each group (non-native teacher, native teacher, and student) were compared. To answer research question 2, the level of variation of the 22 students' translated texts (carried out in random order) was examined using an AI and plagiarism detector, copyleaks.com. For research question 3, the posttest results (including the number of attempts and correct answers) were analysed using descriptive statistics and t-tests. Finally, the Likert-scale responses and follow-up explanations of the post-survey were analysed and compared to the posttest results.

3. Results

The first survey results (see Tables 1, 2, and 3) showed that the EFL learners were more inclined to strongly agree or agree that MT was an effective learning tool; 78% (109/139) compared to the non-native English teachers at 20% (2/10), or the native English teachers at 40% (4/10). Concerning whether MT was better than a dictionary for learning, 40% (4/10) of the native English teachers strongly agreed or agreed, compared to 20% (2/10) of the non-native English teachers and 20% (28/139) of the EFL learners. As to whether student use of MT was unavoidable, 50% (5/10) of the non-native teachers agreed compared to 80% (8/10) of their native teacher counterparts. Overall, the majority of teachers (native and non-native) felt that MT was more beneficial for higher level learners (see Table 4). However, a large percentage of native teachers (40% (4/10)) also stated that MT could be useful for beginners or learners of all levels.

Judging from the feedback, many of the EFL learners considered MT as an effective learning tool for its capacity to quickly and accurately contextualise words and phrases, and easily facilitate wholistic understanding. Although somewhat apprehensive, the native English teachers' responses showed that they are more willing than their non-native counterparts to embrace the technology. One native teacher hypothesised that "less reliance on comprehension and memorisation" and more focus "on having the students do something with what is being taught" could help promote more acceptance of its use. Another native teacher stated that, "The sooner we embrace the technology in the classroom the sooner we can share ideas and communicate our thoughts." However, some of the non-native teachers' concerns were that "MT is a practical tool, not a learning tool" and over reliance on it may cause students to "abandon trying to think in English."

Table 1. EFL learner perceptions of MT (N=139)

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
1. MT is a good tool for learning	25	84	14	15	0

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
2. MT is more effective than a dictionary for learning	5	23	49	52	10

Table 2. Non-native English teachers' perceptions of MT (N=10)

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
1. MT is a good tool for learning	1	1	3	4	1
2. MT is more effective than a dictionary for learning	0	2	3	4	1
3. Using MT in the classroom is good	1	2	2	1	4
4. Using MT for assignments (homework) is good	0	3	2	3	2
5. MT is unavoidable, L2 teachers must accept it	1	4	1	2	2

Table 3. Native English teachers' perceptions of MT (N=10)

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
1. MT is a good tool for learning	0	4	1	5	0
2. MT is more effective than a dictionary for learning	1	3	2	4	0
3. Using MT in the classroom is good	1	3	2	2	2
4. Using MT for assignments (homework) is good	1	2	2	4	1
5. MT is unavoidable, L2 teachers must accept it	4	4	2	0	0

Table 4. MT is an effective learning tool for which students? (N=20)

	Advanced learners	Intermediate learners	Beginners	All learners	No learners
Native English Teachers ' Responses	3	3	1	3	0
Non-Native English Teachers' Responses	6	2	2	0	0
Overall Responses	9	5	3	3	0

Analysis of the 44 translated paragraphs using the AI and plagiarism detector (Appendix C) revealed that the DeepL-assisted ones were either deemed to be 100% identical (paragraph one) or 65.5% identical, with the other 34.5% containing only minor variations (paragraph two). For the paragraphs translated with an online dictionary, 100% of paragraph one and 71.7% of paragraph two were identified as paraphrased versions of each other. The remaining 28.3% were found to be more similar but not identical, with only minor differences. The learner-produced online dictionary translations exhibited more language variation than DeepL, suggesting that more individual effort was invested in their construction. Although the original texts were both three sentences long, the average English translation for paragraph one was 65 words, compared to 55 words for paragraph two.

The posttest results showed that a substantially higher number of attempts were made to translate the key words of each paragraph when the students used the online dictionary. DeepL only promoted a 67% attempt rate and a 36% success rate. The online dictionary promoted a 93% attempt rate and a 65% success rate. The *t*-Test analysis revealed that significant differences existed in these results (Attempts - $p < .01$ / Success Rate - $p < .01$). The post-survey results found that a higher percentage of the EFL learners preferred using DeepL to the online dictionary to do the translation tasks (55% vs. 45%). However, in correlation with the posttest data, 100% of them felt that the online dictionary promoted more language awareness and 95% stated it was more effective for language learning.

For the EFL learners, using the online dictionary lead to more hesitation and indecision about the suitability of their grammar. One student commented that it took him "a long time to make three sentences" and that he "was not sure if the meaning [would be] understood." On the other hand, another student commented that DeepL helped her "know correct sentence structures and expressions," while another stated it was "so easy and quick…" Essentially, DeepL facilitated fewer opportunities for the learners to reflect on their language output. Due to its ability to generate high-quality translations, most of the learners lacked confidence or motivation to limit their use of it as a reference tool for learning.

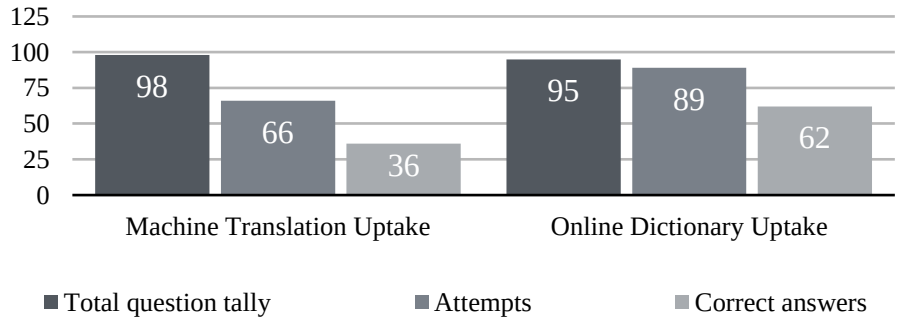


Figure 1. Participants' Posttest Resul

4. Discussion

For research question 1, the results showed that the EFL learners were more inclined than the teachers to believe that MT is a beneficial tool for language learning. Compared to a conventional dictionary, its capacity to deliver precise translations at the click of a button was seen as a notable advantage. The non-native English teachers held the most reservations about its potential, with only 20% stating that it could facilitate learning or that it was more effective than a dictionary. A common theme in the non-native teachers' feedback was that MT is not conducive to language acquisition as that is not its designed purpose. However, its capacity to generate translations that facilitate discussions about acquired knowledge might benefit learners, depending on the nature of the class. Some native English teachers also held reservations about MT's use for language learning. However, there was generally a stronger belief in this group that student use of MT needed to be accepted or even embraced. Willingness to adopt its use was seen as an important first step in creating dialogue about how to proactively manage its use to benefit learners.

For research question 2, analysis of the translated texts found a greater variation in the EFL students' vocabulary and sentence structures when the online dictionary was used than DeepL. Most of the DeepL translations were found to be identical or almost identical in nature. For paragraph two, the smaller word count likely increased the percentage of online dictionary translations (28.3%) categorised as being more similar than paraphrases. For DeepL, the segments identified as having only minor variations consistently matched two specific points of the text, the use of the word '*but'* or '*however'* and the phrase '*way of learning'* or '*learning method'*. This suggests that these parts were more likely to have been generated randomly by DeepL than by the learners themselves. During the task, it was observed that most students finished their translations more quickly using DeepL, with only two students being observed to have limited their use of it, by only translating words or phrases or by spending time rewording the DeepL-translated version afterwards.

Finally, for research question 3, language awareness during the translation task was promoted more by the online dictionary than DeepL. The efficacy of MT to deliver precise, comprehensive, and context-specific translations resulted in the EFL students spending less time considering particular vocabulary or phrases. It is possible that the majority felt the complexity of the MT-generated sentence structures were more advanced than what they could produce on their own, resulting in an over-reliance on the technology and diminished opportunities for learning to occur. If the learners' proficiency levels had been higher, it is possible that more of them would have attempted to translate the texts in their own words and limited DeepL use for reference purposes. Creating opportunities for the learners to discuss the MT-translated texts may also have helped increase metalinguistic awareness.

5. Conclusions

There is growing evidence that MT is increasingly being used by L2 students to assist their language learning, particularly for writing. As MT technology continues to improve, it will likely become even more prevalent in academic EFL contexts. Using DeepL can offer learners the chance to recognise language patterns, connections between the structure and significance of language, word choices, and patterns of collocation. However, relying too heavily on MT and taking shortcuts may not be helpful for individuals who aim at developing advanced writing skills in foreign languages. Consequently, it is important to investigate the effects its use may have on L2 development. The findings of this study showed that exposure to translated output by DeepL did facilitate language awareness. However, it did not do so to the same extent as the online dictionary. The initial survey results found that most of the EFL learners felt MT was a good learning tool. However, when provided with an opportunity to use DeepL to translate a task in a classroom environment, most of the learners who participated ended up producing little of their own output. To promote autonomy and self-directed learning strategies that can help learners become successful writers, it is vital to put them in situations where they must actively engage in the writing process. It is not possible to prevent MT use outside of the classroom. Therefore, it is important to provide training and opportunities that teach learners how to use MT more appropriately and responsibly.

References

Baraniello, V., Degano, C., Laura, L., Lozano, M., Zahonero, M., & Petroni, S. (2016). A wiki-based approach to computer-assisted translation for collaborative language learning. In Y. Li, M. Chang, M. Kravcik, E. Popescu, R. Huang, Kinshuk, & S. Chen (Eds.), *State-of-the-art and future directions of smart learning: Lecture notes in educational technology* (pp. 369–379). Singapore: Springer. https://doi.org/10.1007/978-981-287-868-7_45

Godwin-Jones, R. (2015). Contributing, creating, curating: Digital literacies for language learners. *Language Learning & Technology*, *19*(3), 8–20. http://dx.doi.org/10125/44427

Innes, A. R. B. (2019). Differentiating between translation and student translation: Red flags salient lexicogrammatical features. *Lublin Studies in Modern Languages and Literature*, *43*(4), 1-13. http://dx.doi.org/10.17951/lsmll.2019.43.4.1-13

Jolley, Jason R. & Luciane Maimone (2022). Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. *L2 Journal*, *14*(1): 26-44. http://repositories.cdlib.org/uccllt/l2/vol14/iss1/art2

Wong, T., & Lee, J. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, *28*(2), 187–206. https://doi.org/10.1017/S0958344015000257

Appendix A

- Teacher Survey

1. Online translation software is an effective learning tool students should use to study English.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

2. Please provide a reason for your thinking.

3. It is more effective for students to use online translation software than a dictionary to improve their English.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

4. Please provide a reason for your thinking.

5. I am happy for students to use online translation software in the classroom.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

6. Please provide a reason for your thinking.

7. I am happy for students to use online translation software to complete assignments (homework included) outside of the classroom.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

8. Please provide a reason for your thinking.

9. Which type of learners do you feel online translation software benefits the most?

Advanced learners

Intermediate learners

Beginners

All learners

No learners

10. Please provide a reason for your thinking.

11. I feel online translation software is unavoidable, so it needs to be embraced by L2 teachers.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

12. Please provide a reason for your thinking.

- Student Survey

1. What grade are you in?

1st year

2nd year

3rd grade

4th grade

Other:

2. How would you rate your English level?

Advanced

Upper-Intermediate

Intermediate

Early Intermediate

Elementary

Other:

3. What translation software do you use most frequently?

DeepL

Google Translate

Microsoft Translate

Weblio

Other: 4.

4. How often do you use translation software?

Every day

Weekly

Every month

Not so much

Never

Other:

5. How did you first learn about translation software?

6. How do you mainly use translation software?

To translate one word at a time

To translate phrases one at a time

To translate one document at a time

To translate one paragraph at a time

To translate a large number of sentences

Other:

7. How often do you use translation software for English assignments (including homework) outside of class?

All the time

Quite a bit

Sometimes

Not so much

Never

Other: 8.

8. Do you think translation software is a good tool for learning English?

Yes, very much

I don't know

Not really

Not at all

Other: 9.

9. Please write your reason

10. Do you think using translation software rather than a dictionary to improve your English is better?

Yes, very much

Yes, very much

I don't know

Not really

Not at all

Other: 11.

Please write your reason.

Appendix B

Translation Task: Paragraph one

第二言語習得とは、人間が母語以外の第二言語を習得するプロセスを科学的に解明する学問のことを指します。そのプロセスには言語学だけではなく脳科学や心理学、社会学など幅広い領域が関わってくる。第二言語の習得プロセスが分かれば、そのプロセスに沿った適切な英語学習トレーニングを積んでいくことで、効率的に英語力を高めることができます。

[Second language acquisition refers to the scientific study of the process by which humans acquire a second language other than their native language. The process involves not only linguistics, but also brain science, psychology, sociology, and a wide range of other fields. Once the process of second language acquisition is understood, English language skills can be efficiently improved by building on appropriate English language learning training that follows the process.]

Translation Task: Paragraph two

第二言語習得について一定の知識を持っておくことは、英語学習の効率を高めるうえでとても役立つのです。しかし第二言語習得自体は特定の学習方法ではなく、あくまで言語習得プロセスに関する理

論となります。そのため、英語学習を行う上では、実際にその理論を具体的な学習方法に落とし込む必要があります。

[Having a certain knowledge about second language acquisition is very helpful in increasing the efficiency of English language learning. However, second language acquisition itself is not a specific learning method, but rather a theory of the language acquisition process. Therefore, it is necessary to actually put the theory into concrete learning methods when learning English.]

Post-Test: Translate Japanese to English

- From paragraph one

解明する [study, investigate, enquire, explore, understand, elucidate, clarify]

社会学 [sociology]

脳科学 [brain science, neuroscience]

関わる [engage, involve, relate to]

積む [accumulate, build]

- From paragraph two

特定, [particular, specific, certain]

理論, [theory]

具体的, [specific, concrete, particular]

落とし込む [put into practise, put into action, apply]

Post Survey

1. For this translation task, which did you prefer to use: DeepL or the online dictionary?

DeepL

Online dictionary

2. Please write down your reasons

3. Which one made you think more deeply about language?

DeepL

Online dictionary

4. Please write down your reasons

5. Which one helped you learn more English?

DeepL

Online dictionary

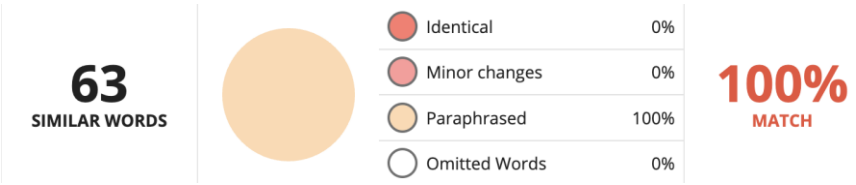
6. Please write down your reasons

Appendix C

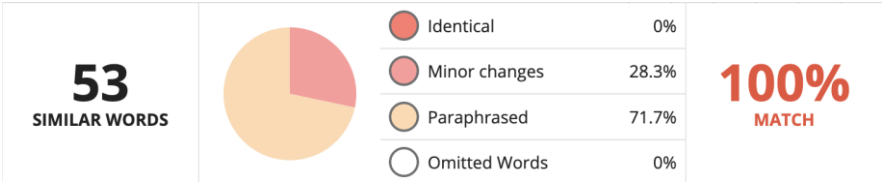
DeepL: Paragraph One

DeepL: Paragraph Two

Online Dictionary: Paragraph One



Online Dictionary: Paragraph Two



EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16912

Support inclusive teaching and learning with CALL technologies in the beginning-level Chinese language classroom

Lisha Xu[a]
[a]Asian Studies, Mount Holyoke College, , lxu@mtholyoke.edu

How to cite: Xu, L. (2023). Support inclusive teaching and learning with CALL technologies in the beginning-level language classroom. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16912

Abstract

The importance of creating inclusive language learning environments in foreign language classrooms has gained recognition, driven by the diverse learning styles and the needs of learners (Jones, 2013). However, providing effective support to all learners remains a substantial challenge due to the inherent diversity among learners and the prevalence of various learning difficulties.In the context of Chinese as a Second/Foreign Language (CSL/CFL), research and practice have predominantly emphasized instructional methods targeting content-related challenges, often overlooking the development of inclusive language learning environments tailored to the diverse needs of learners (Tao et al., 2021). This study, therefore, investigates the pedagogical approaches and digital technology implementations within a specific course to assess their effectiveness in promoting inclusivity in the classroom. The study also assesses the alignment of these strategies with the unique characteristics and requirements of CSL/CFL teaching and learning.To gather insights, the researcher conducted a comprehensive analysis of course evaluations and administered a post-course survey to gauge student perceptions. The findings indicate that the implemented strategies effectively promoted inclusiveness and facilitated Mandarin learning in the first-year Chinese classroom.

Keywords: *Inclusive foreign language classroom, CSL/CFL, Pedagogy, CALL.*

1. Introduction

The review of research on inclusive education in CSL/CFL sheds light on the lack of explicit pedagogical practices and research aimed at fostering a more inclusive classroom environment to address diverse needs (Tao et, al., 2021). Carter (2021) suggests that this dearth of practices can be attributed to the unique aspects of CSL/CFL content, classroom dynamics, and broader educational contexts. Surveys conducted by Carter (2021) regarding inclusive pedagogical practices in CFL classrooms argue that inclusion may manifest differently in the CFL context. Mandarin, being an orthographic language with a tonal system, is widely regarded as challenging for English speaker (Tao et, al., 2021). Moreover, with the growing interest in Mandarin learning, learners have become increasingly diverse, presenting additional challenges for CSL/CFL practices. This leads us to the question: How can inclusive education methods be rendered more effective?

Universal Design for Learning (UDL) (CAST, 2018) proposes that inclusivity should encompass affective, recognitional, and strategic aspects of learning. It highlights the significance of enhancing learners' motivation, employing multimodal approaches to present knowledge, and fostering engagement with instructors and learning materials (Carter, 2021). Abbott (2013) affirms that digital technologies can aid, support, and empower language learning, making them valuable tools for including all learners in a multilingual future. Carter (2021) further

advocates for the use of digital tools to provide diverse supports that enhance inclusivity and assist practitioners in designing activities to address challenges in CSL/CFL classrooms. Therefore, transitioning to the UDL approach and utilizing technologies to facilitate its implementation are viable paths toward achieving inclusive CSL/CFL instruction and addressing the distinctive teaching and learning obstacles associated with Mandarin.

This study examines inclusive practices and strategies supported by digital tools, with a focus on addressing the following research question: To what extent do learners perceive the effectiveness of utilizing these methods to enhance inclusiveness and facilitate Mandarin learning in the first-year Chinese classroom?

2. Method

This study utilizes an action research approach to assess the effectiveness of pedagogical approaches. Data for the study was collected through an anonymous online questionnaire administered to students at the conclusion of the Spring 2023 semester. The questionnaire was conducted using Google Forms, which facilitated the collection of responses. The questionnaire consisted of 26 questions, all in English to ensure clarity and avoid potential misunderstandings. It included six short-answer questions and 20 multiple-choice questions. The questions were divided into three sections:

1. Participants' Chinese learning experience and learning styles.

2. Participants' experience and attitude towards the strategies employed to build a learning community and enhance proficiency in the four language skills.

3. Participants' perception of the effectiveness of the technology tools used to accommodate various learning styles and assist with Chinese character learning.

The development of inclusive teaching methods and survey questions drew upon relevant studies by Jones (2013), Alexander (2008), Jiang (2017), Ji (2017), Lü (2017), Mou (2003), CAST (2018), Shen & Liao (2017), Zhang (2021), Cornelius (2013), Han, Liu, Sun (2023), and Higbee (2009).

Additionally, insights from the course evaluation provided by the college were taken into consideration. These insights covered course materials, workload, and perspectives on teaching methods that promote effective learning and cater to individual needs.

2.1 Instructional Context and Participants

This study focused on the curricular design implemented by the researcher for two semesters, beginning with Chinese language courses, AS110 and AS111, during the academic year of 2022-2023. These courses were offered at a private college in the US. Each course consisted of 15 weeks per semester, with a typical week consisting of three 50-minute lectures and two 50-minute oral practice sessions. The majority of learners in the courses were native English speakers from the US, while some were heritage learners or had prior learning experience before joining the class. Placement tests were administered to determine the appropriate course level for these students. It is worth noting that there were students who submitted an accommodation letter indicating their dyslexia condition. For further demographic data, please refer to Table 1 below.

Table 1. Demographic data

	AS110(Total 22 students)	AS111(Total 14 students)
Language system	English:18; Others (Nepali, Vietnamese):3; Logographic (Japanese): 1	English:12; Others (Nepali, Vietnamese): 2
Heritage learners	4	2
Learned Chinese before	10	7

Table 1. Demographic data

	AS110(Total 22 students)	AS111(Total 14 students)
Dyslexic	2	1

2.2 Data collection

Course evaluations for AS 110 and AS 111 were collected at the conclusion of the Fall 2022 and Spring 2023 semesters, respectively. Out of the 22 students enrolled in AS 110, 15 submitted course evaluations. Similarly, out of the 13 students in AS 111, five submitted course evaluations. Additionally, five students from AS 111 responded to the questionnaire designed by the researcher. While the sample size in this study is small, it is important to note that the participants represent a diverse range of language learning experiences and styles, making them representative of different groups of CFL/CSL learners in US college-level Chinese programs.

3. Results and discussion

3.1 The findings from the course evaluation offered valuable insights

The software Nvivo was used to categorize learners' comments regarding the course materials, tools, and teaching methods used to facilitate learning and promote inclusive education. The positive comments mainly highlighted aspects related to accommodating learners' needs and utilizing a variety of materials and technological tools. This was evident in a representative selection of students' responses. Please see the Table 2 below.

Table 2. Students' responses to the course evaluation

Student	Response
1	"The various learning materials she uses also help students with various learning styles grasp new information and concepts."
2	"I did study a lot outside of class which helped a lot. I used the slide shows, quizlets, and kahoots that the professors made for us."
3	"She used a wide variety of course materials including vocabulary quizzes, workbook assignments, extensive readings, video projects, lesson quizzes, and writing assignments on Miro. All of these materials helped improve my Chinese proficiency and allowed me to practice all aspects of the language."
4	"She incorporated a variety of learning tools and was receptive to student feedback about what worked and what didn't. Specific resource examples include online programs, workbook assignments, and traditional in-class tests and quizzes."

In the course evaluation report from AS111, a student also mentioned that the tutoring sessions with a native speaker for the real-life speaking practice project "were definitely helpful in learning the language, but we did it so often that a lot of students were tired with it from time to time." This project aimed to assist students with different proficiency levels in engaging in language practice, facilitating real-world interactions, and providing various avenues for expression at their own pace. The practice also offered flexible self-assessment opportunities for students to test their abilities. To further explore the benefits and limitations of this project, future interviews should be conducted to gain a better understanding of the methods and frequency of this assignment.

3.2 The results from the questionnaire provided practical ideas

The questionnaire was distributed to the participants of the AS111 course in mid-May 2023, and a total of five (38%) responses were collected. Among the five students, two had prior learning experiences before joining the class, and one spoke Mandarin or a dialect at home. The survey results pertaining to pedagogical practices for

building a learning community, enhancing real-world interactions, and accommodating diverse learning styles and needs are presented below in Table 3. The results were rated on a five point Likert scale, ranging from "Strongly Disagree" to "Strongly Agree."

Table 3. Students' responses to the survey

Survey question	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The methods used in classroom teaching helped to build a learning community that is collective, reciprocal, supportive, cumulative, and purposeful.	4		1		
The tutoring session helps to meet your learning needs.	3	2			
The tech tools used in and out of class provide multiple ways to engage in learning.	2	3			
The course provides multiple means of assessing knowledge, which help you develop the four language skills.	1	4			
The different types of projects increase your motivation to learn and cater to your personal learning needs and styles.	1	4			

As evident from the aforementioned responses, the outcomes were affirmative concerning the strategies employed to enhance inclusivity and support Mandarin learning in the first-year Chinese classroom. In summary, inclusive CFL pedagogical practices encompass several key facets.

Firstly, the curriculum design should focus on making the learning process relevant, engaging, and rewarding, as emphasized by Carter (2021), and providing support for personalized learning. This approach helps to motivate learners and keep them on track. The feedback from some course participants further highlights the effectiveness of this approach:

1. "At the end of this course, I was able to carry conversations about daily life with a native Chinese speaker."
2. "I felt like I was retaining a lot of information while completing course work."
3. "I have found myself able to understand snippets of Chinese conversation I hear in day-to-day life, and I especially am proud of my ability to read and write in characters."

These responses indicate that the curriculum design, which focused on relevance, engagement, and personalization, has positively impacted the learners' language proficiency and skills. The ability to engage in conversations, retain information effectively, and comprehend real-life Chinese interactions demonstrates the success of the inclusive pedagogical approach. Furthermore, the learners' pride in their achievements in reading and writing Chinese characters highlights the effectiveness of the curriculum in addressing the challenging aspect of CFL learning.

Secondly, materials should be presented in multiple modes to cater to different learning styles. For instance, providing audio for vocabulary quizzes helps support vocabulary learning. Chinese character learning, known for its difficulty and time-consuming nature, benefits from the inclusive approach facilitated by technological tools. These tools provide inclusive methods for character recognition, retention, and production. Students found tasks that combined handwriting and typing to be efficient in alleviating anxiety during character learning, thus creating an inclusive language learning environment.

Thirdly, inclusive learning should provide multiple channels for expression and offer choices in certain aspects of learning. For example, students can choose their topics and ways of presenting their projects. In terms of assessment, incorporating self-assessment and informal assessment practices enhances learning. One such

practice is in-class quick writing through typing, where students answer questions or share personal experiences related to the topics learned. This practice, commonly known as 'exit ticket', helps instructors gauge students' understanding of the materials (Dixson & Worrel, 2016; Cornelius, 2013; Carter, 2021).

4. Conclusions

This paper reflects on the pedagogical practices of inclusive education at the beginning level of the CFL classroom, using course evaluations and a specific questionnaire as the survey instrument. The findings suggest that, at least for this small sample, the methods were effective in fostering an inclusive classroom. However, the instructor still faces various challenges in making individual accommodations and designing differentiated instruction, particularly in the acquisition of Chinese characters at the beginning level and for students with learning disabilities. Additionally, while incorporating technological tools to support learning proved effective, the instructor should be mindful of limiting the number of assignments and the variety of platforms used as there is a possibility of overwhelming learners, as indicated in a response to the questionnaire. Lastly, further research is needed on providing inclusive learning materials specifically tailored to the beginning level.

References

Abbott, C. (2013). Technology Uses and Language – A Personal View. In E. Vilar Beltrán, C. Abbott & J. Jones (Ed.), *Inclusive Language Education and Digital Technology* (pp.30-44). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847699749-004

Alexander, R. (Ed.) (2008). *Culture, dialogue and learning: Notes on an emerging pedagogy*. SAGE Publications Ltd. https://doi.org/10.4135/9781446279526

Cornelius, K. E. (2013). Formative Assessment Made Easy: Templates for Collecting Daily Data in Inclusive Classrooms. *TEACHING Exceptional Children*, 45(5), 14-21. https://doi.org/10.1177/004005991304500502

CAST (2018). Universal Design for Learning Guidelines version 2.2. http://udlguidelines.cast.org

Carter, Liz. (2021). What does the post-secondary Mandarin foreign language inclusive classroom look like? The state of the art and some practical suggestions. *International Journal of Chinese Language Education*, 9, 17-58.

Dante D. Dixson & Frank C. Worrell (2016). Formative and Summative Assessment in the Classroom, Theory Into Practice, 55:2, 153-159. https://doi.org/10.1080/00405841.2016.1148989

Higbee, J. L. (2009). Implementing Universal Instructional Design in Postsecondary Courses and Curricula. *Journal of College Teaching & Learning (TLC)*, 6(8). https://doi.org/10.19030/tlc.v6i8.1116

Han, J., Liu, Q., & Sun, R. (2023). A Multimodal Approach to Teaching Chinese as a Foreign Language (CFL) in the Digital World. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 13(1), 1-16. http://doi.org/10.4018/IJCALLT.322026

Jones, J. (2013). Modern Foreign Languages as an Inclusive Learning Opportunity: Changing Policies, Practices and Identities in the Languages Classroom. In E. Vilar Beltrán, C. Abbott & J. Jones (Ed.), *Inclusive Language Education and Digital Technology* (pp. 3-29). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847699749-003

Jiang, Xin (2017). Teaching Chinese Characters in the Keyboard Era. *International Chinese Language Education*, 2017(2), 4-10.

Ji, Honggang (2017). Implications of Character Processing Research for Chinese Reading Instruction. *International Chinese Language Education*, 2017(2), 10-17.

Lü, Chan (2017). The Acquisition of Chinese Characters and Words by Heritage Learners of Chinese in the United States. *International Chinese Language Education*, 2017(2), 21-26.

Mou, Sherry J. (2003). Integrating Writing into Elementary Chinese. *Journal of the Chinese Language Teachers Association*, 38(2), 109-136.

Shen, H. & Liao, J. (2017). Towards Digitalization: Research on Teaching Chinese as a Second Language Vocabulary and Characters in the 21st Century. *International Chinese Language Education*, 2017(2), 26-33.

Tao, Hongyin & Carter, Liz & Wan, Helen & Zhou, Yan. (2021). Inclusive Education in Chinese as a Second/Foreign Language: An Overview and Research Agenda. 9. 1-15. *International Journal of Chinese Language Education*, 9, 1-15.

Zhang, P. (2021). Typing to Replace Handwriting: Effectiveness of the Typing-Primary Approach for L2 Chinese Beginners. *Journal of Technology and Chinese Language Teaching*, 12(2), 1-28.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16996

The use of machine learning in developing learner-adaptive tools for second language acquisition

Maryam Sadat Mirzaei[a] and Kourosh Meshgi[b]
[a]RIKEN Center for Advanced Intelligent Project, maryam.mirzaei@riken.jp and [b]RIKEN Center for Advanced Intelligent Project, kourosh.meshgi@riken.jp

How to cite: Mirzaei, M.S.; Meshgi, K. (2023). The use of machine learning in developing learner-adaptive tools for second language acquisition. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16996

Abstract

Advancements in artificial intelligence and machine learning present opportunities to revolutionize language learning tools with learner-adaptive capabilities. These technologies facilitate the creation of trainable systems that can interact with learners, offering personalized learning experiences tailored to individual needs, interests, proficiency levels, backgrounds, and native languages. This study explores the role of machine learning in developing personalized frameworks for second language learning, introducing the Partial and Synchronized Caption (PSC) tool as an example. PSC utilizes automatic speech recognition and natural language processing to identify challenging words for language learners, which are presented in the caption while masking easy words. We used machine learning to personalize the caption for various learners. An experiment involving graduate students learning English as a second language demonstrated the adaptability of PSC's word selection to different learners. While creating entirely personalized captions may be challenging, PSC offers a promising approach to personalized and adaptable listening tools. The data collected from learner interactions also provides valuable insights into individual needs, shaping future language learning tools and pedagogical practices.

Keywords: *Machine learning, learner-adaptive technologies, personalized language learning, Partial and Synchronized Caption.*

1. Introduction

Machine learning has enabled the production of personalized technologies that focus on individual users' preferences and needs, emerging in the form of recommendation systems, personal AI assistants, news feeds, and many more. Recently, leveraging data-driven algorithms has enabled tailoring learning experiences by providing personalized, custom-made learning practices for individual learners based on their unique needs, interests, and proficiency levels, especially in the domain of second language learning (Chen et al., 2021). The main goal is to deliver: (a) precise and appropriate content; (b) accurate assessments; (c) tailored feedback; and (d) a customized learning path. In this context, many studies have focused on the development of learner-centric tools and content personalization to address diverse learners' preferences. These studies cover a range of applications, including personalized readability assessment for L2 reading (Ehara, 2022), difficulty detection and adaptation for practicing grammar (Pandarova et al., 2019), personalized conversational AI agents (Dizon et al., 2022), robot-assisted language learning (Randall, 2019), and simulation and games (Karoui et al., 2021; Peterson & Jabbari, 2022). Notably, there is a specific focus on personalized mobile-assisted learning (Gumbheer et al., 2022).

The development of personalized profiles or learner models facilitates effective content filtering, directing learners toward resources that match their specific proficiency levels, learning objectives, and content preferences (Godwin-Jones, 2017). Personalization occurs at different levels. Traditionally adaptive systems focused on coarse-grained personalization, providing general and broad-level personalization by drawing on user demographics and grouping users into broader categories based on shared characteristics, preferences, or behaviors (Walkington & Bernacki, 2020). The advantages include easier implementation and reduced data requirements within these broad categories. However, more complex adaptation is needed to provide accurate and tailored learning practice that is aligned with individual learners' needs (Ismail et al., 2016). The fine-grained personalized system addresses these limitations by considering the unique preferences and behavior of each user, accounting for specific interactions, historical data, and individual feedback to provide a highly personalized learning experience. Nonetheless, incorporating learner diversity requires more individual user data. Furthermore, personalization can be combined with in situ learning to provide personalized content that is contextually and situationally relevant (Dalton-Puffer & Smit, 2013). Finally, dynamic and progressive personalization includes the continuous update of the system based on the learner's progress and evolving needs for long-term practice and learning outcomes (Shute & Rahimi, 2017).

To provide an example, we introduce a tool for L2 listening development called Partial and Synchronized Caption (PSC). With this method, we aim to move toward a learner-centered listening practice, with the ability to fine-tune the system based on the learner's preferences and advancement constantly.

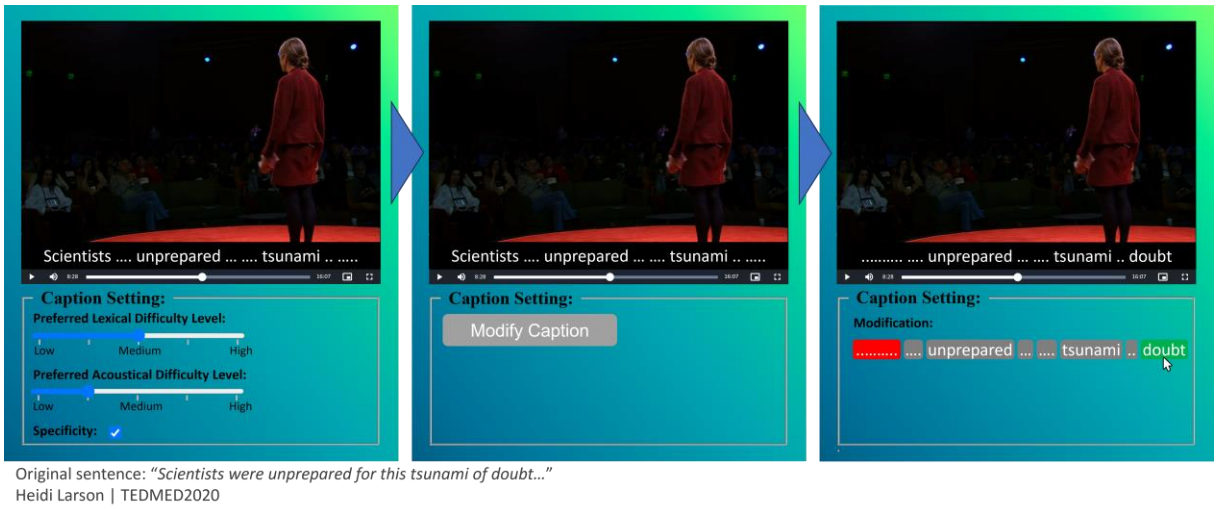


Figure 1. PSC self-regulation feature (left) and word choice personalization feature (middle and right)

2. Personalized caption

PSC uses lexical and acoustic features to determine the difficulty level of the words and decides which parts of the text need to be presented in the caption, intending to encourage listening over reading. The caption is synchronized on the word level, thereby enabling seamless audio and text mapping and reducing distraction.

For the baseline version, we used rule-based coarse-grained level assignments to roughly categorize learners into three language proficiency levels (beginners, intermediate, advanced) based on learner's assessment tests (TOEFL/TOEIC score, speech rate tolerance, vocabulary size). Word selection is determined by defining thresholds for specific features, including word frequency and speech rate, while also incorporating additional factors like automatic speech recognition system errors, word specificity, proper names, and abbreviations (Mirzaei et al., 2018). Compared to the full caption and keyword caption, the baseline version of PSC provides a certain level of personalization by adapting the words in the caption to different proficiency levels. Yet, within each group, learner variability poses a challenge, requiring further adaptation to meet individual user preferences

and requirements. Therefore, we employed two distinct approaches that integrate two modes of learner feedback (Figure 1).

2.1. Self-paced learning through learner customization

Self-directed learning is crucial, emphasizing individuals' initiative in diagnosing needs, adapting strategies, and evaluating outcomes. For listening, self-paced learning personalizes the experience, fostering agency, and motivation, and reducing cognitive overload, leading to improved performance (Ozcelik et al., 2019). Acknowledging that learners may have varying strengths in vocabulary retention and tolerance for fast speech, we have integrated a user-friendly interface, where users can easily modify the system's parameters and adjust the number of words displayed, hence promoting autonomous and self-regulated learning. This level of customization allows learners to tailor the language learning process to their unique preferences, pace, and proficiency. Those with a robust vocabulary can opt for a fewer number of words to challenge themselves further. On the other hand, individuals who struggle with fast speech can increase the number of words with a faster speech rate displayed, allowing them to focus on mastering the content at a pace that suits their comfort level. The flexibility of the learning environment encourages learners to explore and experiment freely.

2.2. Personalization through learner feedback

This subsequent stage involves automatic personalization or adaptation. During this phase, we gather learners' feedback to enhance the word choices within the caption and align it with their preferences. If a learner opts to modify the word selection, the words within the caption become visible in clickable boxes. At this point, if a learner finds a particular word unnecessary in the caption, that word hides and is marked for subsequent system retraining. Conversely, if a learner wishes to reveal a hidden word, a simple click unveils it, prompting the system to update accordingly.

To achieve adaptation, the system computes the differences between the original generated caption and any modifications introduced to it. Following this, the system adjusts its parameters to optimally align with the marked words by dynamically fine-tuning the coefficients of the extracted features, utilizing supervised learning. This iterative process ensures that the system continuously improves its personalized word selection based on user interactions and feedback. The classifier is retrained with this data, allowing it to grasp insights into each individual's language learning challenges, background, vocabulary knowledge, and potential sources of listening difficulties to enhance personalized captions.

3. Experiments

We conducted a preliminary experiment with this tool involving 29 participants who were intermediate English learners. They were graduate students, aged 22-26, from diverse academic backgrounds including fields such as medicine, business, and engineering. The material used for the experiment consisted of TED talks delivered by native English speakers.

Participants accessed the experimental session online. Ahead of the experiment, they acquainted themselves with the system by viewing three brief videos and testing the self-regulation (adjusting default parameters) and feedback collection (editing shown/hidden words in the caption) functionalities. After this introductory episode, the first stage of the experiment involved watching four videos (V1-V4) with baseline PSC, where participants could customize the system parameters for self-regulated listening. They adjusted the frequency and speech rate threshold for the generated caption during the first five minutes of each video, creating their desired settings. Whenever they faced listening difficulty, they could pause the video and use the clickable box to see the hidden words they couldn't recognize. They could also omit unnecessary/distracting words. The system stored this user feedback to fine-tune its future word selection and generate captions that aligned more with user preferences. After each video, participants received a report detailing the percentage of shown/hidden words in their personalized caption, motivating them to remove unwanted words to provide room for desired ones.

In the second stage of the experiment, the system used learner data from the previous videos to better adapt its word selection for the subsequent four videos (V5-V8). Participants continued to use the modification function to show/hide words when necessary. The learners' log files from both stages were stored for further analysis. We also had a brief interview with learners to capture any feedback or suggestions.

4. Results and discussion

Figure 2 illustrates learner feedback on the system, analyzed through log files across the two stages of the experiment, using distinct videos. In the initial stage, participants actively engaged with the system, refining captions to align with their proficiency and demands. The feedback collection feature allowed for pinpointing challenging words and eliminating unnecessary ones, resulting in more personalized captions. Notably, interaction peaked at this point, suggesting active customization of captions to match learners' expectations. The focus was primarily on shown words, likely due to their impact on listening difficulties. However, as depicted in the figure, the volume of learner modifications decreased as the experiment progressed to the next stage.

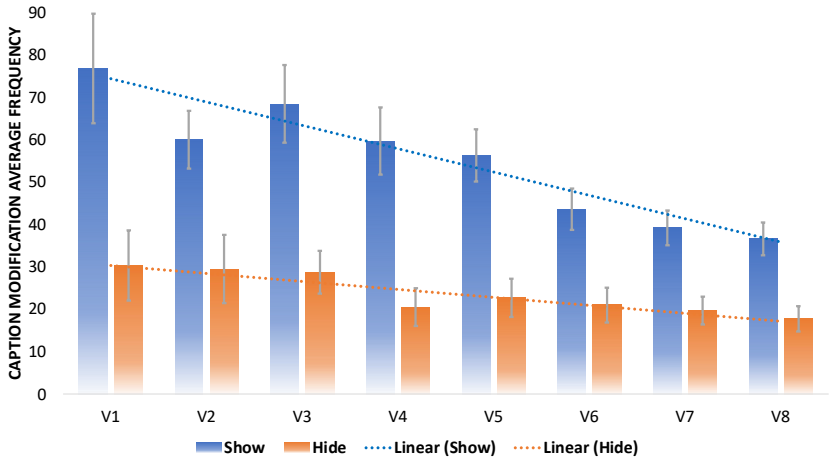


Figure 2. Experimental result on learners' modification of PSC's word choices

The second stage showcased the PSC system's adaptability (V5-V8). Leveraging data from the initial stage, the system further improved its word selection for subsequent videos, enhancing overall personalization. Fewer modifications in this stage underscores the PSC's ability to adapt its word selection, capturing essential words for individual learners to be shown in the caption. Nevertheless, there remained potential for improvement by omitting unnecessary words based on learner feedback.

We discuss two primary factors in Figure 2. One pertains to **learner variability** while the other is associated with **content variability**. The experiment's outcome highlights differences in show/hide word feedback across our learners. The observed deviation (STD bar in the figure) in the number of modifications made by learners highlights learner variability, illustrating that some needed more adjustments while others required fewer. This emphasizes the diversity among learners within the same group.

When examining the videos themselves, it becomes apparent that, for certain videos, modifications increased (V3&V5), and most learners required an increased number of shown words. This observation aligns with the learners' feedback during the interview phase where most learners indicated that these videos were notably challenging, featuring either fast speech rate or technical language, and subject complexity. However, two learners did not express particular difficulty with these videos as they shared a background related to the video's topic. This finding is important, as it introduces another factor (learner background) that should be considered in the context of adaptation.

The interview data also indicated that certain learners utilized the modification feature to reveal hidden words, to confirm their correct recognition, after which they opted to hide the word again. Conversely, this also implies that

learners might hide some words under the assumption of their ease of recognition, potentially influenced by having seen those words in the caption. This suggests that learners might not always have a comprehensive awareness of their requirements for self-regulation. Additionally, learner feedback provided valuable insights into interface improvements, the inclusion of repetition, and the ability to select preferred videos.

5. Conclusions

While creating a caption that is truly designed for each learner may not be entirely feasible, we found that PSC could adapt its word selection to an acceptable level for different learners. This presents a promising avenue for crafting personalized, adaptable listening tools using machine learning, automatic speech recognition, and natural language processing. Notably, learner and content variations significantly impact personalization and system efficacy. It's worth noting that if adaptable technology falls short of user expectations, it can detrimentally affect satisfaction and engagement. Finally, our results underscore the need to expand findings across various proficiency levels and backgrounds as well as compare with other existing methods, such as keyword captioning in future research. Additionally, tailoring feedback and enabling reflection and self-monitoring are essential considerations.

References

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning. *Educational Technology & Society, 24*(1), 205-222.

Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: a research agenda. *Language Teaching, 46*(4), 545-559. https://doi.org/10.1017/s0261444813000256

Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers and Education: Artificial Intelligence, 3*, 100088. https://doi.org/10.1016/j.caeai.2022.100088

Ehara, Y. (2022). Uncertainty-aware Personalized Readability Assessment Framework for second language learners. *Journal of Information Processing, 30*, 352-360. https://doi.org/10.2197/ipsjjip.30.352

Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. *Language Learning & Technology, 21*(1), 4–15. https://dx.doi.org/10125/44592

Gumbheer, C. P., Khedo, K. K., & Bungaleea, A. (2022). Personalized and adaptive context-aware mobile learning: review, challenges and future directions. *Education and Information Technologies, 27*(6), 7491-7517. https://doi.org/10.1007/s10639-022-10942-8

Ismail, H. M., Harous, S., & Belkhouche, B. (2016). Review of personalized language learning systems. In *2016 12th International Conference on Innovations in Information Technology* (pp. 1-6). IEEE. https://doi.org/10.1109/innovations.2016.7880051

Karoui, A., Alvarez, L., Goffre, T., Dherbey Chapuis, N., Rodi, M., & Ramalho, M. (2021). Adaptive pathways within the European platform for personalized language learning PEAPL. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 90-94). https://doi.org/10.1145/3450614.3464480

Mirzaei, M. S., Meshgi, K., & Kawahara, T. (2018). Exploiting automatic speech recognition errors to enhance partial and synchronized caption for facilitating second language listening. *Computer Speech & Language*, *49*, 17-36. https://doi.org/10.1016/j.csl.2017.11.001

Ozcelik, H. N., Van den Branden, K., & Van Steendam, E. (2019). Listening comprehension problems of FL learners in a peer interactive, self-regulated listening task. *International Journal of Listening*, 1-14. https://doi.org/10.1080/10904018.2019.1659141

Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R. D., & Brefeld, U. (2019). Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education, 29*, 342-367. https://doi.org/10.1007/s40593-019-00180-4

Peterson, M., & Jabbari, N., eds. (2022). *Digital games in language learning: case studies and applications.* Taylor & Francis. https://doi.org/10.4324/9781003240075-1

Randall, N. (2019). A survey of robot-assisted language learning (RALL). *ACM Transactions on Human-Robot Interaction (THRI), 9*(1), 1-36. https://doi.org/10.1145/3345506

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*(1), 1-19. https://doi.org/10.1111/jcal.12172

Walkington, C., & Bernacki, M. L. (2020). Appraising research on personalized learning: definitions, theoretical alignment, advancements, and future directions. *Journal of research on technology in education, 52*(3), 235-252. https://doi.org/10.1080/15391523.2020.1747757

EUROCALL 2023: CALL for All Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16994

Using generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic

Annika Simonsen[a] and Branislav Bédi[b]
[a]University of Iceland, ans72@hi.is and [b]Árni Magnússon Institute for Icelandic Studies, branislav.bedi@arnastofnun.is

How to cite: Simonsen, A.; Bédi, B. (2023). Using Generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16994

Abstract

We utilize ChatGPT and other generative AI tools in developing illustrated multimodal resources for learning Icelandic as a second language (L2) via reading in the online platform LARA. These are illustrated short stories in Icelandic created using specific prompts including information about learner levels and age groups. We present a methodology for creating engaging stories and a way to evaluate the quality and suitability of text and images generated by AI. Additionally, we assess how 46 adult learners perceived reading a story at A1 and A2 level in LARA, with a positive vocabulary learning effect. We conclude that generating suitable prompts greatly assists with getting the desired output; however, this is restricted to the language one works with. Much human post-editing is still necessary for Icelandic texts to improve their quality in grammar, vocabulary, and cultural aspects, and their suitability for language teaching.

Keywords: *generative artificial intelligence, ChatGPT, prompt engineering, Icelandic L2, language education, reading.*

1. Introduction

Current generative Artificial Intelligence (AI) tools assist with rapid and relatively reliable ways of generating text and images. Despite the significant strength of ChatGPT-4 in generating language content for English, its capability is currently limited when it comes to less commonly spoken languages (Chang et al., 2023) including Icelandic. Our aim is to find out whether ChatGPT-4 can be used to generate text sutitable for language teaching, post-reading exercises, and illustrations to assist L2 Icelandic learners with enhancing their reading comprehension skills. Using ChatGPT-4 to create educational resources for less resourced languages such as Icelandic can be time and cost efficient. Here, we focus on creating multimodal resources, specifically reading texts sutitable for language teaching, post-reading exercises, and illustrations to assist L2 Icelandic learners with enhancing their reading skills. These resoures will be presented in a multimodal fashion in the LARA (Learning And Reading Assistant) platform, which enables recordings of individual words and text segments through the Text-To-Speech (TTS) technology of Google and presenting translation of words into English (in our case) and a concordance page with references to frequency and alphabetical lists.

Through ChatGPT-4's Natural Language Processing (NLP) capabilities, we aim to provide L2 learners of Icelandic with an engaging language learning environment. To achieve this, the content must be multimodal and educational (Dressmann, 2019), i.e. combining various senses including reading, listening, illustrations, and

exercises. We utilize different AI system for creating text (ChatGPT-4) and images (DALL-E-2[1], Midjourney[2] and Stable Diffusion[3]), and the LARA platform for integrating the illustrated reading texts.

We present results from an anonymous online survey, which was conducted about reading perception. Responses from 46 adult L2 Icelandic learners from fourteen different countries with age range 18-64 years, at beginner (A1) to intermediate (A2) levels, were obtained. The following section will describe the method for generating text and images. The tesults are discussed in the final section.

2. Method

Constructing effective prompts is crucial in getting relevant responses from ChatGPT (Kohnke et al., 2023). We developed specific prompts that helped create desired content. The ChatGPT capabilities enabled creating lemmatised versions of texts with translations into English and short stories for different learner levels based on the Common European Framework of Reference for Languages (CEFR).[4] This was done thanks to ChatGPT-4's capability to align texts with the CEFR framework as well as manually incorporating this information into our prompts. These texts were furthermore alighed with the corresponding overview of reading comprehension guidelines (Piccardo, 2020) for age groups (pupils in primary schools, students in secondary schools, and adults in colleges). Texts and images were subsequently transformed into multimodal resources in the online platform LARA (Akhlaghi et al., 2019).

Additionally, inspired by Eldan & Li (2023), who created a synthetic dataset of English short stories *TinyStories* generated by GPT-3.5 and GPT-4 that contain only words that three to four year olds would understand, we used only part 1 and 3 of the following framework: 1) a list of words to be included in the story, 2) a sentence that should appear somewhere in the story, 3) a list of features (possible features: dialogue, bad ending, moral value, plot twist, foreshadowing, conflict), and 4) a short summary (one to two lines) of the story. The authors used GPT-4 to evaluate its own stories by asking to act as a teacher and grade the stories as if they were written by a student. Our prompts were generated in two languages, English and Icelandic, to test the ChatGPT capability in generating the same content using two different languages.

2.1. Generating texts and exercises

Following the OpenAI guidelines[5] for generating prompts, we focused on parameters that would be specifically relevant to teachers of L2 Icelandic. We manually incorporated information about CEFR, the reading comprehension guidelines (Piccardo, 2020), into our prompts, and specified whether the readers are children or adults. Additionally, we requested ChatGPT to provide multiple-choice questions for the stories based on the difficulty level of the target learner audience. We also provided a list of specific words that needed to be included in the stories to simulate teachers' purpose to possibly teach specific words in their class. Finally, we tried writing prompts in both English and Icelandic, both resulting in similar content. After ChatGPT had generated the stories from the prompts, they were proofread by human native speakers. Firstly, by a teacher of L2 Icelandic for adults, who proofread fifteen stories (see Appendix A for the prompts used to create the stories). The stories included content which was often not suitable to the cultural setting of Iceland, such as restaurant staff calling restaurant guests by their first names. Secondly, selected stories were revised by additional human evaluators to reflect a real-world Icelandic cultural setting (e.g. adjusting the names of persons and places, plot points in the local setting, and the description of places to create a more authentic sounding environment). Through the editing phase, some aspects of the story, such as the vocabulary, setting, and cultural context were changed to make sense of the plot in the story.

[1] https://openai.com/dall-e-2

[2] https://www.midjourney.com

[3] https://stablediffusionweb.com

[4] https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

[5] https://platform.openai.com/docs/guides/gpt-best-practices

Table 1. Overview of the experimental setup and parameters based on 15 stories (see Appendix B for costs associated with generating stories with ChatGPT-4).

Experimental setup	Parameters
Stories without any given subject	Length: 250 words Learner level: adult beginner
Stories with a given subject	Topic: "Write a short anecdotal story in Icelandic about buying an ice cream in Reykjavík"
Stories with different CEFR levels	CEFR Levels: A1, A2, B1, B2, C1, C2
Stories for children	Age: 3-5, 6-7, 8-10, 11-13, 14-17 years old
Stories with specific features and words	Feature list (e.g. "plot twist, dialogue, hero") and word list (e.g. "cat", "sitting", "happy")
Icelandic prompts	Prompt language: Icelandic
Creation of multiple-choice questions	Output type: multiple choice questions
Self-evaluation of stories	Role: Teacher evaluation
Grammatical proofreading by native speaker	Role: Native Icelandic speaker, L2 teacher
Real-world revisions	Revisions: Names, places, plot points

For regarding exercises, we specified the number of items in multiple choice questions, the learner level, and that they had to be written in Icelandic and include correct answers.

2.2. Generating images

We used DALL-E 2, Midjourney, Stable Diffusiont's demo version on HuggingFace[6], and DreamStudio[7] to generate images to illustrate the stories. Prompts were written by both the human content creator and ChatGPT-4. The most successful results were obtained by showing ChatGPT-4 a prompt guide[8] written by the developers of Stable Diffusion and asking it to write an illustration creation prompt for a specific story. This prompt was then used to generate an image with either DALL-E 2, Midjourney, or Stable Diffusion. The results varied, and it quickly became evident that the different AI systems require different amounts of detail in their prompts. In the end, one image was chosen from the selection provided by very different tools and art styles (compare Figure 1, 2, and 3).

[6] https://huggingface.co/spaces/stabilityai/stable-diffusion

[7] https://beta.dreamstudio.ai/generate

[8] https://stable-diffusion-art.com/prompt-guide/

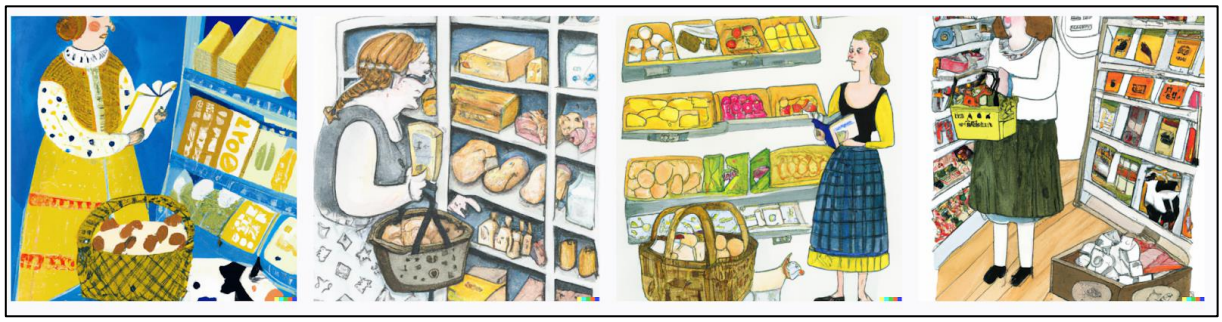


Figure 1. Illustration generated by DALL-E 2 using prompt, “A whimsical book illustration of a woman shopping in an Icelandic grocery store with a shopping basket containing potatoes, meat, and cheese. She's looking at the milk section, preparing to select the next item”.

Figure 2. Illustration generated by Stable Diffusion using prompt, “A book illustration of a woman shopping in an Icelandic grocery store with a shopping basket containing potatoes, meat, and cheese”.

Figure 3. Illustration generated by Midjourney using prompt, “Digital painting of a woman named Hildur shopping in an Icelandic grocery store. She is holding a shopping basket filled with potatoes, meat, and cheese. In front of her is the milk section, which she is carefully considering. The scene is vibrant and detailed, in the style of a contemporary Icelandic artist”.

3. Assessing the reading experience of L2 Icelandic learners

Two out of 15 illustrated short stories (one at A1[9] and one at A2[10] level) were selected for evaluation by 46 adult L2 Icelandic learners from 14 different countries with age 18-64 years, either at beginner to below intermediate (26 learners at A1 level) or intermediate (20 learners at A2 level). An anonymous online survey was made using Google Forms with 21 multiple choice, closed and open-ended questions about the story, learning experience, perception of AI presence, and demographics[11], and distributed on social media and by email in an ongoing summer language course. Link to the online version of the two stories in LARA was provided (see example in Figure 4).

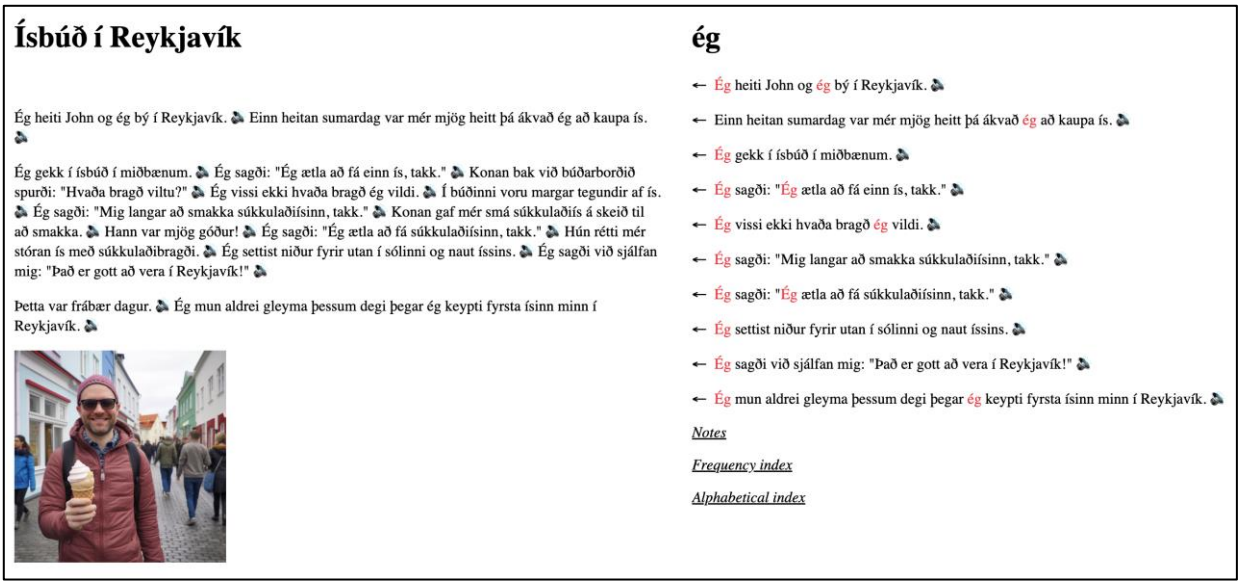


Figure 4. The A1 story shared with participants on the online platform LARA, with illustration created by Midjourney (right) and concordance page (left).

4. Results and discussion

The results indicate a practical benefit in using generative AI tools for creating unique content suitable for language learning at different learner levels and age groups. ChatGPT-4, however, still makes many grammatical errors in Icelandic and the prompts need to include specific instructions about the desired length of texts. Human post-editing is the most time-consuming part (ca. 25 minutes per story). By comparison, the grammatical Word Error Rate (WER) in our stories was relatively high (22-25%), likely due to the fact that the proofreader was a teacher and was asked to make the text suitable for language learning. The other two proofreaders did not receive such instructions and their WER was lower (12.80-20%). The WER does not include cultural and contextual inconsistences.

ChatGPT-4 is, nonetheless, able to correct some of its own grammatical errors in Icelandic, e.g. grammatical gender and tense, but not consistently enough (see Appendix C). Although ChatGPT-4 seems less capable of correcting nonsensical words and phrasing, it can produce different styles based on learner age, e.g. a story about a woman going grocery shopping includes the incorrect term in Icelandic *húðflúrsfylling* (tattoo-filling) instead of *smjördeigsfylling* (puff-pastry filling). Stories can also contain peculiarities, e.g. a tourist from New York visiting Iceland named *Ólafur* (typical Icelandic name) instead of having a typical US name, or a man buying an

[9] https://www.issco.unige.ch/en/research/projects/callector/1044_Ísbúð_í_Reykjavíkvocabpages/_hyperlinked_text_.html

[10] https://www.issco.unige.ch/en/research/projects/callector/1040_Daniel_frá_Winnipegvocabpages/_hyperlinked_text_.html

[11] https://github.com/BranBedi/ChatGPT-LARA/blob/ea7524f8417dfc4d2f28563746f6c00453def08c/Questionnaire%20for%20A1%20and%20A2%20stories%20in%20LARA.pdf

ice cream (not a drink) because of being thirsty. Human editors need to add more cultural aspects, e.g. description of town architecture, interior of typical cafés, or appropriate phrases for ordering a coffee. These would make the stories more engaging. We observed that ChatGPT-4 does indeed have the capability of adding jokes to stories but, at this stage, it would more often than not make nonsensical or inappropriate jokes in Icelandic.

For adult learners, ChatGPT-4 typically created stories about people of foreign origine living in Iceland and learning Icelandic in different settings. For children learners, the typical story would involve children or heroic animals. To get more variety in the plots, using Eldan & Li's (2023) framework was helpful. Reading guidelines and the CEFR reference helped ChatGPT-4 to make the story adjusted to the learner level.

To generate images, we recommend to show ChatGPT a prompt guide[12] and ask it to write a prompt for an image in each story. Various generative AI tools can help create different illustrations that are suitable for different styles and reader audience, and/or personal taste of the content creator.

For exercises, ChatGPT-4 is good in creating simple and effective multiple choice questions. If specified, it can also generate answers that can be proofread by the content creators themselves.

In the online survey, L2 Icelandic learners of both reading groups, A1 and A2 stories, reported high reading comprehension, 32 out of 46 (70%) reported learning new vocabulary (one or more new words)[13] from the stories, but commented that they could have appreciated more engaging content and more cultural representation of Iceland, e.g. eating ice cream during winter instead of summer. Suggestions for future iterations included enhancing the user experience by creating more content similar to the stories used here, however, focusing on detailed, culturally-rich narratives to facilitate a more immersive learning experience, and improving compatibility with reading these texts on mobile devices. When asked how likely it was that the story had been generated by AI, the A1 reading group showed slightly more belief in this idea. This might be because the A2 story went through more post-editing process by a human.

By combining the capabilities of AI with human expertise during the post-editing stage, our goal is to attain a balance between efficiency and ethical content creation.

5. Conclusions

In conclusion, the ChatGPT-4 shows promising potential for generating text sutitable for L2 Icelandic learners to help enhance reading skills. However, human expertise is necessary during post-editing to include cultural accuracy. ChatGPT-4 still makes many grammatical and lexical errors when generating text in Icelandic and lacks local cultural knowledge to give it appropriate setting and language context because narratives need to be both linguistically accurate and culturally resonant. Future endeavors should also focus on enriching the stories with cultural information and possibly amusing content to foster a more immersive learning experience.

Acknowledgements

We thank Ásdís Helga Jóhannesdóttir for proofreading the stories, Manny Rayner for his invaluable support in this project, to all participants involved in this experiment, and to the Icelandic Student Research Innovation Fund for funding this research.

References

Akhlaghi, E., Bédi, B., Butterweck, M., Chua, C., Gerlach, J., Habibi, H., Ikeda, J., Rayner, E., Sestigiani, S., &

[12] https://stable-diffusion-art.com/prompt-guide/

[13] The English translations on the LARA platform do not work on smartphones, so several survey respondents were unable to access them.

Zuckermann, G.(2019). In *Proceedings of 8th ISCA Workshop on Speech and Language Technology in Education (SLaTE2019)*, 99-103. https://www.isca speech.org/archive/pdfs/slate_2019/akhlaghi19b_slate.pdf.

Bédi, B., ChatGPT-4, Chiera, B., Chua, C., Chiaráin, N. N., Rayner, M., Simonsen, A & Zviel-Girshin, R. (2023). ChatGPT + LARA = C-LARA. ResearchGate. https://www.researchgate.net/publication/373952306_ChatGPT_LARA_C-LARA.

Chang, Y., Wang, X., Wang, J., Wu, Y., Zhu, K., Chen H., Yang, L., Yi, X., Wang, C., Wang. Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q. and Xie, X. (2023). A survey on evaluation of large language models. In *J. ACM, 37,* (4), Article 111. Published online. https://arxiv.org/pdf/2307.03109v8.pdf.

Dressman, M. (2019). Multimodality and Language Learning. In M. Dressman and R.W. Sadler (Eds), *The Handbook of Informal Language Learning* (pp. 39-55). https://doi.org/10.1002/9781119472384.ch3.

Eldan, R., & Li, Y. (2023). TinyStories: How Small Can Language Models Be and Still Speak Coherent English?. In *arXiv*. Published online. https://arxiv.org/pdf/2305.07759.pdf.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868.

Piccardo, E. (2020). The Common European Framework of Reference (CEFR) in language education: Past, present, and future. *Monterey, CA & Baltimore, MD: TIRF & Laureate International Universities*. https://www.tirfonline.org/wp-content/uploads/2021/10/LEiR_CEFR.pdf.

Appendices

Appendix A

List of prompts

1. **A1-level short story with a topic provided**
   a) You are a teacher teaching Icelandic as a second language to adults. Write a short anecdotal story (approximately 250 words) in Icelandic about buying an ice cream in Reykjavík. Make sure the text is appropriate for an adult L2 learner of Icelandic who is a beginner.

   b) Excellent. Now change the text, so that it is appropriate for an adult L2 learner of Icelandic who is intermediate.

   c) Excellent. Now change the text, so that it is appropriate for an adult L2 learner of Icelandic who is advanced/master.

2. **A2-level short story without a topic provided**
   a) You are a teacher teaching Icelandic as a second language to adults. Write a short anecdotal story (approximately 250 words) in Icelandic. Make sure the text is appropriate for an adult L2 learner of Icelandic who is a beginner.

   b) Excellent. Now change the text, so that it is appropriate for an adult L2 learner of Icelandic who is intermediate.

   c) Excellent. Now change the text, so that it is appropriate for an adult L2 learner of Icelandic who is advanced/master.

d) Thank you. What is the difference between the first, second and third story you made in terms of difficulty?

3. Showing ChatGPT the overall reading comprehension guidelines from CEFR

a) This is an overview of Overall reading comprehension from the Common European Framework for Languages. A1-A2 is beginners, B1-B2 is for intermediates and C1-C2 is for advanced.

C2: Can understand virtually all forms of the written language including abstract, structurally complex, or highly colloquial literary and non-literary writings. Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.

C1: Can understand in detail lengthy, complex texts, whether or not they relate to his/her own area of speciality, provided he/she can reread difficult sections. Can understand a wide variety of texts including literary writings, newspaper or magazine articles, and specialised academic or professional publications, provided that there are opportunities for re-reading and he/she has access to reference tools.

B2: Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.

B1: Can read straightforward factual texts on subjects related to his/her field and interests with a satisfactory level of comprehension.

A2: Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or jobrelated language. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.

A1: Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.

Pre-A1 Can recognise familiar words accompanied such as a fast-food restaurant menu illustrated with photos or a picture book using familiar vocabulary.

You are a teacher teaching Icelandic as a second language to adults. Write a short anecdotal story (approximately 250 words) in Icelandic. Make sure the text is appropriate for an adult L2 learner of Icelandic who is level A1 or A2 according to the Common European Framework for Languages.

b) Great. Now change the story so it's appropriate for an adult L2 learner at level B1 to B2.

c) Great. Now change the story so it's appropriate for an adult L2 learner at level C1 to C2.

d) Thank you. Could you tell me what the difference is between the three stories in terms of difficulty?

Making multiple choice questions with ChatGPT

e) Could you make five multiple choice questions for the first story? They should be appropriate for adult L2 learners at level A1 to A2.

f) And could you now make five multiple choice questions for the second story? They should be appropriate for adult L2 learners at level B1 to B2.

g) And could you now make five multiple choice questions for the third story? They should be appropriate for adult L2 learners at level C1 to C2.

h) Could you give the answers to all of the multiple-choice questions you just made?

4. Showing ChatGPT-4 examples of LARA stories before asking for a story

a) LARA (Learning and Reading Assistant) is a collaborative open-source project, active since mid-2018, whose goal is to develop tools that support conversion of plain texts into an interactive multimedia form designed to support development of L2 language skills by reading. The basic approach is in line with Krashen's influential Theory of Input, suggesting that language learning proceeds most successfully when learners are presented with interesting and comprehensible L2 material in a low-anxiety situation.

Here are a few examples of LARA stories created by ChatGPT:

Dutch

Oma wint skateboardwedstrijd in Nederlands dorp In het kleine Nederlandse dorpje Rolbord, heeft een 75-jarige oma, genaamd Jannie, iedereen verbaasd door de lokale skateboardwedstrijd te winnen. Jannie, een voormalig gymnastieklerares, heeft laten zien dat leeftijd slechts een getal is en dat je nooit te oud bent om te leren en plezier te hebben.

Jannie, die pas een jaar geleden begon met skateboarden, had al snel de aandacht getrokken van de dorpsbewoners. Ze oefende dagelijks in het lokale skatepark, waar ze nieuwe trucjes leerde en vriendschappen sloot met jongeren uit het dorp. Haar doorzettingsvermogen en enthousiasme waren aanstekelijk, en haar populariteit groeide snel.

Toen de jaarlijkse skateboardwedstrijd naderde, besloot Jannie zich in te schrijven en deel te nemen. Hoewel sommigen twijfelden aan haar vermogen om te concurreren met de jongere deelnemers, was Jannie vastberaden om iedereen te bewijzen dat ze het kon.

Op de dag van de wedstrijd wist Jannie het publiek te verbazen met haar indrukwekkende vaardigheden en behendigheid op het skateboard. Ze voerde ingewikkelde trucjes uit, zoals de "ollie" en de "kickflip", en won uiteindelijk de eerste prijs.

De winst van Jannie heeft niet alleen geleid tot veel felicitaties van de dorpsbewoners, maar heeft ook de aandacht getrokken van nationale en internationale media. Haar verhaal heeft velen geïnspireerd en laten zien dat het nooit te laat is om nieuwe uitdagingen aan te gaan.

Nu haar overwinning bekend is geworden, heeft Jannie plannen om haar vaardigheden verder te ontwikkelen en zelfs skateboardinglessen te geven aan andere senioren in Rolbord. Zo hoopt ze anderen te inspireren om uit hun comfortzone te stappen en het beste uit zichzelf te halen, ongeacht hun leeftijd.

Icelandic:

Páfagaukur reddar tónleikum Reykjavík

– Í gær kvöld voru tónleikar á skemmtistaðnum Hljóðbylgja, þar sem hljómsveitin Fjallabrjót tók áhorfendur á óvart með óvæntan gest. Á meðal laganna sem þeir fluttu var páfagaukur, sem hafði skellt sér inn í tónleikasalinn. Áhorfendur voru fyrst undrandi en brátt fylgdust þeir með áhuga. Páfagaukurinn flaug um salinn, skellti sér upp á svið og fylgdi takti tónlistarinnar með blíðu hávaða. Hljómsveitin tók þetta með jólum og spilaði saman við fuglinn. Áhorfendur klappuðu og öskruðu þegar tónleikarnir laukust, en það var engin vafi um að páfagaukurinn hefði stolið sýninguna. Eftir tónleikana var fuglinum fylgt út, en hann hélt áfram að syngja með Fjallabrjótum, sem stóðu þar úti og sungu nokkur lög að skemmti fólks. Þessi sérstaki tónleikakvöld munu vafalaust verða mikið rædd í komandi dögum, og páfagaukurinn hefur nú orðið algjör staðarfrægur.

Italian:

Un gatto diventa il sindaco di un piccolo paese italiano In un piccolo paese italiano chiamato Gattopoli, i cittadini hanno votato a sorpresa un gatto di nome Fuffi come il loro nuovo sindaco. Sembra che Fuffi, un felino molto amato nella comunità, abbia conquistato i cuori degli abitanti grazie al suo fascino e alla sua presenza costante nella piazza del paese.

Fuffi, un gatto nero con una macchia bianca sul petto, è noto per aver aiutato a risolvere piccoli problemi tra i cittadini e per aver tenuto lontani i piccioni dal centro storico. Anche se il suo curriculum politico potrebbe sembrare limitato, i cittadini di Gattopoli hanno deciso di affidargli la guida del paese.

Durante la campagna elettorale, Fuffi ha ricevuto il sostegno di molti residenti, che hanno creato cartelloni e slogan per promuovere il loro candidato felino. "Fuffi per un Gattopoli migliore" e "Una zampa avanti con Fuffi" sono solo alcuni degli slogan che hanno riempito le strade del paese.

Il risultato delle elezioni ha attirato l'attenzione dei media internazionali, che sono accorsi a Gattopoli per intervistare Fuffi e i suoi sostenitori. Il vice sindaco, Luigi Rossi, ha assicurato che Fuffi sarà affiancato da un team di esperti umani che lo aiuteranno a prendere decisioni importanti per il bene del paese.

Nonostante le ovvie difficoltà pratiche, i cittadini di Gattopoli sono entusiasti della loro scelta e sperano che Fuffi porti un'atmosfera di armonia e collaborazione nel paese. Dopotutto, come dicono gli abitanti del posto, "Fuffi sa come mettere d'accordo tutti".

You are a teacher teaching Icelandic as a second language to adults. Write a short anecdotal story (approximately 250 words) in Icelandic that is in the style of the examples shown before. Make sure the text is appropriate for an adult L2 learner of Icelandic who is a beginner.

Asking for CEFR-related changes

b) Change the story so that it's more in line with the reading comprehrension guidelines from The Common European Framework of Reference for Languages for beginners levels A1 and A2.

A2: Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or jobrelated language. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.

A1: Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.

Asking ChatGPT to stay within the word limit

c) Make sure the stories are about 250 words long while still being in line with the reading comprehension guidelines from The Common European Framework of Reference for Languages for beginners levels A1 and A2.

5. **Asking for a story for children by providing ages**

    a) You are a teacher teaching Icelandic as a second language to children. Write a short anecdotal story (approximately 250 words) in Icelandic that is in the style of the examples shown before. Make sure the text is appropriate for a child L2 learner of Icelandic who is between 3-5 years old.

    b) Excellent, now make a story that is appropriate for an child L2 learner of Icelandic who is between 6-7 years old.

    c) Excellent, now make a story that is appropriate for an child L2 learner of Icelandic who is between 8-10 years old.

    d) Excellent, now make a story that is appropriate for an child L2 learner of Icelandic who is between 11-13 years old.

    e) Excellent, now make a story that is appropriate for an child L2 learner of Icelandic who is between 14-17 years old.

    f) Thank you. Could you tell me what the difference is between the three stories in terms of difficulty?

6. **Asking for a story for children by providing CEFR levels**

    a) You are a teacher teaching Icelandic as a second language to children. Write a short anecdotal story (approximately 250 words) in Icelandic that is in the style of the examples shown before. Make sure the text is appropriate for a child L2 learner of Icelandic who is level A1.

    b) Excellent, now make a story that is appropriate for a child L2 learner of Icelandic who is level A2.

    c) Excellent, now make a story that is appropriate for a child L2 learner of Icelandic who is level B1.

    d) Excellent, now make a story that is appropriate for a child L2 learner of Icelandic who is level B2.

    e) Excellent, now make a story that is appropriate for a child L2 learner of Icelandic who is level C1.

f) Excellent, now make a story that is appropriate for a child L2 learner of Icelandic who is level C2.

g) Thank you. Could you tell me what the difference is between the three stories in terms of difficulty?

7. Asking for a story featuring specific words and features (inspired by Eldan and Li (2023).

a) You are a teacher teaching Icelandic as a second language to adults. Write a short anecdotal story (approximately 250 words) in Icelandic. The story should include the following features: plot twist, dialogue, hero, and the words "cat", "sitting" and "happy". Make sure the text is appropriate for for an adult L2 learner of Icelandic who is a beginner.

8. Following Eldan and Li (2023)'s framework for evaluating stories

a) Complete the Icelandic story following the symbol ***.

b) You are a teacher teaching adults Icelandic as a second language. In the following exercise, the student is given a beginning of a story. The student needs to complete it into a full story. The exercise tests the student´s language abilities and creativity. The symbol *** marks the separator between the prescribed beginning and the student's completion:

Einu sinni, í fornri húsnæði, bjó stúlka sem hét Lily. Hún elskaði að skreyta herbergið sitt með fögrum hlutum. Eitt kvöld fann hún stóran kassa á vindinum. Hún opnaði hann og sá marga glitrandi skreytingar. Lily var mjög glöð og ákveðaði að nota þær í herberginu sínu. Þegar Lily var að skreyta herbergið sitt, myrkvaði loftið út um gluggana. Það var hávaðasamur*** stormur sem hófst, og vindurinn blés kaldan andanum inn í herbergið. Þrátt fyrir það, held Lily áfram að skreyta. Hún var mjög einbeitt og lét enga óróleika trufla sig. Hun hengdi glitrandi skreytingarnar um allt í herberginu. Þær skínu sem stjörnur í myrkri. Þegar hún var loks búin að skreyta allt, sat hún niður og horfði á verk sitt. Það var eins og að horfa upp í stjörnuþokuna, hver skreyting var sem stjarna sem skín í nóttinni. Skyndilega, í miðri storminum, heyrdi hún banka á dyrnar. Hún stóð upp og fór að opna. Þar stóð maður með grátt hár og skyggndist inn í herbergið. Hann brosti þegar hann sá hvernig Lily hafði skreytt herbergið. "Það er fallegt, Lily," sagði hann. "Þú hefur skapað þitt eigið alheimur hér inní." Stormurinn lagðist smám saman og maðurinn fór aftur út í nóttina. Lily horfði eftir honum og brosti. Hún vissi að hún hafði skapað eitthvað sérstakt. Herbergi hennar var ekki bara herbergi lengur, það var alheimur fullur af stjörnum. Og það var allt saman takk sé einum glitrandi skreytingum sem hún hafði fundið í kassanum á vindinum.

c) Please provide your general assessment about the part written by the student (the one after the *** symbol). Is it grammatically correct? Is it consistent with the beginning of the story? Pay special attention to whether the student manages to complete the sentence which is split in the middle by the separator ***.

d) Now, grade the student's completion in terms of grammar, creativity, consistency with the story's beginning and whether the plot makes sense.

9. Asking ChatGPT to write a prompt for DALL-E 2, Midjourney and Stable Diffusion.

a) Make a DALL-E 2 text prompt for the first story you made.

b) Make a Midjourney text prompt for the first story you made.

c) Make a Stable Diffusion text prompt for the first story you made.

d) Here is a prompt guide from Stable Diffusion. Read it and make a prompt for the first story you made. [prompt guide inserted]

Appendix B

Overview over basic requirements to follow our methodology.

Step	Requirement	Cost	Notes
1	Subscription to ChatGPTplus	$20 per month	The alternative is to apply for API access for GPT-4. With API-access to GPT-4, you can use the playground to generate texts using system prompting. The user is charged for input and output tokens.
2	Formulating prompt	N/A	ChatGPT-4 will always recommend that the user writes the prompt in English, but it is also possible to write it in Icelandic.
3	A native speaker to correct the output	Variable	If the user is not a native speaker, then the cost will depend on the specific rates of the native speaker that they hire.
4	Registration with DALL-E 2, Midjourney, or Stable Diffusion	$15 per 115 prompts for DALL-E 2 tokens or $10 per month for Midjourney subscription	DALL-E 2 users from before April 6th 2023 get free credits. Stable Diffusion is free, but requires a good computer and IT knowledge for the best result. Midjourney requires a Discord account.

Appendix C

This appendix contains an example that shows the difference between using a professional proof-reader to proofread an Icelandic story generated by ChatGPT-4 (figure 5) and using ChatGPT-4 to proofread its own story (figure 6). The professional proof-reader is a native speaker of Icelandic, who teaches L2 Icelandic to adults. She was told to proofread the text, so it was appropriate as teaching material for L2 speakers of Icelandic. ChatGPT-4 was asked to proofread its own story in a new window and was told it was a story written by a student. Additionally, we show two examples of two additional native speakers of Icelandic, who are not teachers, correcting the same text (figure 7 and 9). These proof-readers were not told that this text was meant as teaching material for L2 speakers before proofreading.

Story 1:
A1
Ég heiti John og ég bý í Reykjavík. ~~Eitt~~ sumardag var ég mjög heitur. Ég ákvað**Einn heitan sumardag ákvað ég** að kaupa ís.

Ég gekk ~~til ísbúðar sem er~~í ísbúð í miðbænum. Ég sagði: "Ég ~~vil~~ ætla að fá einn ís, takk."
Konan ~~að baki~~bak við ~~búðarborðsins~~ búðarborðið spurði: "Hvaða bragð viltu?" Ég vissi ekki hvaða bragð ég vildi. Í búðinni voru margar tegundir ~~íss~~af ís. Ég sagði: "~~Ég vil smakka sjokóladeíss~~Mig langar að smakka súkkulaðiísinn, takk."

Konan gaf mér smá ~~sjokóladeís~~ súkkulaðiís á skeið til að smakka. ~~Það var mjög gott~~Hann var mjög góður! Ég sagði: "Ég ~~tek sjokóladeís~~ætla að fá súkkulaðiísinn, takk." Hún ~~gaf~~ rétti mér stóran ís með ~~sjokóladebragði~~súkkulaðibragði.

Ég ~~sat~~ settist niður fyrir utan í sólinni og ~~njót naut~~ íssins. Ég ~~mælti~~ sagði við sjálfan mig: "Það er gott að vera í Reykjavík!"

~~Það~~ Þetta var frábær dagur. Ég mun ~~alltaf muna þennan dag~~aldrei gleyma þessum degi þegar ég keypti fyrsta ísinn minn í Reykjavík.

Figure 5: Example of an Icelandic story created by ChatGPT-4 and grammatically proofread by a teacher who is a native speaker of Icelandic.

Your Icelandic is already very good, but there are a couple of small corrections to be made:

"Ég heiti John og ég bý í Reykjavík. Einn sumardag var ég mjög heitur." should be "Ég heiti John og ég bý í Reykjavík. Einn sumardag var ég mjög heitur." You use 'einn' when referring to 'one day' in Icelandic.

"Ég sat niður utan í sólinni og njót íssins." should be "Ég settist niður úti í sólinni og njótaði íssins." The verb 'að setjast' is reflexive, and 'njóta' should be in the past tense to match the context of the sentence.

"Ég mælti við sjálfan mig: "Það er gott að vera í Reykjavík!"" should be "Ég sagði við sjálfan mig: "Það er gott að vera í Reykjavík!"" The verb 'mæla' is more formal and less commonly used than 'segja'.

Figure 6: Example of ChatGPT-4 proofreading the same Icelandic as in Figure 5. Prompt used: You are a teacher teaching Icelandic as a second language to adults. Please correct the grammatical mistakes in this story written by a student.

A1:
Ég heiti John og ég bý í Reykjavík. Einn sumardag var mér mjög heitt. Ég ákvað að kaupa mér ís.

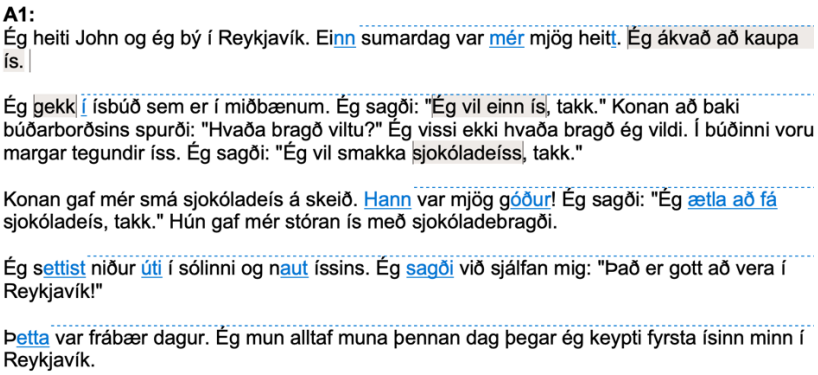
Ég fór í ísbúð sem er í miðbænum. Ég sagði: „Ég ætla að fá einn ís, takk." Konan við búðarborðið spurði: „Hvaða bragð viltu?" Ég vissi ekki hvaða bragð ég vildi. Í búðinni voru margar tegundir íss/ margar tegundir af ís. Ég sagði: „Ég vil gjarna smakka súkkulaðiís , takk."

Konan gaf mér smá súkkulaðiís á skeið. Hann var mjög góður! Ég sagði: "Ég fæ einn súkkulaðiís takk." Hún lét mig hafa stóran ís með súkkulaðibragði.

Ég settist niður í sólinni og naut íssins. Ég sagði við sjálfan mig: „Það er gott að vera í Reykjavík!"

Þetta var frábær dagur. Ég mun alltaf muna þennan dag/ daginn þegar ég keypti fyrsta ísinn minn í Reykjavík.

Figure 7: Example of the A1 Icelandic story created by ChatGPT-4 and grammatically proofread by a non-teacher who is a native speaker of Icelandic.



A1:
Ég heiti John og ég bý í Reykjavík. Einn sumardag var mér mjög heitt. Ég ákvað að kaupa ís.

Ég gekk í ísbúð sem er í miðbænum. Ég sagði: "Ég vil einn ís, takk." Konan að baki búðarborðsins spurði: "Hvaða bragð viltu?" Ég vissi ekki hvaða bragð ég vildi. Í búðinni voru margar tegundir íss. Ég sagði: "Ég vil smakka sjokóladeíss, takk."

Konan gaf mér smá sjokóladeís á skeið. Hann var mjög góður! Ég sagði: "Ég ætla að fá sjokóladeís, takk." Hún gaf mér stóran ís með sjokóladebragði.

Ég settist niður úti í sólinni og naut íssins. Ég sagði við sjálfan mig: "Það er gott að vera í Reykjavík!"

Þetta var frábær dagur. Ég mun alltaf muna þennan dag þegar ég keypti fyrsta ísinn minn í Reykjavík.

Figure 8: Example of the A1 Icelandic story created by ChatGPT-4 and grammatically proofread by an additional non-teacher who is a native speaker of Icelandic.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16948

Virtual reality: “Awesome”, “OK”, or “Not so good” for language learning?

Tricia Thrasher[a], Regina Kaplan-Rakowski[b], Dorothy Chun[c] and Randall Sadler[d]
[a]Immerse, USA, tricia@immerse.online; [b]University of North Texas, USA, Regina.Kaplan-Rakowski@unt.edu; [c]University of California, Santa Barbara, USA, dchun@education.ucsb.edu and [d]University of Illinois at Urbana-Champaign, USA, rsadler@illinois.edu

How to cite: Thrasher, T.; Kaplan-Rakowski, R.; Chun, D.; Sadler, R. (2023). Virtual reality: “Awesome”, “OK”, or “Not so good” for language learning? In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16948

Abstract

High-immersion Virtual Reality (VR) has rapidly gained popularity as an innovative tool that provides users with immersive and engaging learning experiences. Meanwhile, large-scale studies that provide empirical evidence regarding its effectiveness are scarce. This pilot study is part of a Meta and Immerse co-funded large-scale project investigating cognitive and affective aspects of language learning in VR. This short paper reports on one aspect of this pilot: language learners' (n=123) impressions of French as a foreign language (L2) classes conducted in Immerse, a VR language learning platform. Despite its acclaimed potential, VR may not be immediately perceived as an “awesome” tool for language learning. Our findings indicate that most learners viewed VR as "OK" or “awesome” for learning. However, only a few participants expressed VR to be “not so good.” Stemming from positive and negative feedback from the learners, we identify which aspects of VR are most appealing, and which should be carefully considered. Based on lessons learned, we share recommendations of how to avoid typical mistakes when implementing VR in a classroom.

Keywords: *high-immersion virtual reality, language learning, learners’ attitudes, Virtual Reality Assisted Language Learning (VRALL).*

1. Introduction

High-immersion Virtual Reality (VR) has the potential to engage and motivate learners. Scholars in Computer-Assisted Language Learning (CALL) have studied the effectiveness of VR-Assisted Language Learning (VRALL) for practicing social interactions (Dooly et al., 2023; Thrasher, 2022), vocabulary learning (Papin & Kaplan-Rakowski, 2022), virtual exchanges (Gruber et al., 2023), pragmatics (Taguchi, 2022), listening comprehension (Tai & Chen, 2021; Ye & Kaplan-Rakowski, 2024), and reading (Kaplan-Rakowski & Gruber, 2023). Overall, existing research has reported VR to be relatively beneficial for language learning (Dhimolea et al., 2022), however, a common limitation in these studies is the use of small sample sizes, which makes it difficult to generalize findings to larger populations. Our larger project aims to fill this research gap by collecting data on the efficacy of VR in Spanish, French, and English as a Second Language (ESL) classes across 12 different high schools in the USA during the 2023-2024 academic year. As a first part of this large-scale project, a pilot study was conducted testing the cognitive and affective aspects of learning in VR. The goal of this paper is therefore to report on learners' initial views of VR activities after their first three exposures to VR.

Although VRALL has been used and studied for decades and has been found to facilitate language learning (Kaplan-Rakowski, 2011; Lan, 2020; Sadler & Dooly, 2012), most research has involved three-dimensional (3D) virtual worlds, such as *Second Life*, which learners access via two-dimensional displays (e.g. laptops or tablets). This type of VR is considered *low-immersion* compared to *high-immersion* VR which involves the use of VR headsets that allow users to be fully enveloped in a 3D 360-degree environment that they interact with using hand-held controllers. While numerous high-immersion VR apps are available (see *Oculus*[1] store), only a few specialized language learning platforms exist; for instance, *Mondly*, *ImmerseMe*, *Noun Town*, *Language Lab*, and *Immerse*). This paper focuses on research in *Immerse*[2]. This app has been specifically designed for VRALL and developed not only with the fundamental pedagogical principles of learning language in context, having authentic cultural experiences, and developing a vibrant sense of community, but also with capitalizing on some of the unique affordances of VR, namely immersion, presence, and embodiment (Makransky & Petersen, 2021).

A promising theory of immersive learning is Makransky & Petersen's (2021) Cognitive and Affective Model of Immersive Learning (CAMIL). According to this model, a variety of affective and cognitive factors play important roles in how and under what conditions learning occurs or is reinforced. Our paper therefore reports on affective factors, such as interest and agency. However, peripherally, cognitive factors are also considered, including interaction with objects and other learners as well as cognitive load. Based on learners' use of Immerse on three occasions, our research question is: What are language learners' initial opinions about using Immerse for learning French?

2. Method

Participants in the study were 123 (n=123) language learners, who were high-school students of L2 French. The vast majority were from a rural high school in Texas, USA (n_1=109). The remaining learners (n_2=14) were students from a high school in central Illinois, USA. Both schools' learner populations were primarily made up of high-school students from traditionally underrepresented and minority backgrounds. However, a main difference between these two research sites was that the school in Illinois had more financial support since it was part of a larger university system. These two schools were chosen because they had agreed to serve as pilot sites for a larger research project examining the use of VR in 12 high schools throughout the USA.

Learners completed activities in Immerse, which can be accessed via a VR headset or a web browser. For this study, all learners joined sessions using a VR headset. In Immerse, learners can directly interact with their classmates in over 30 different environments (e.g. fast food restaurant, doctor's office, shopping center; Figures 1 & 2). Outside of class time, learners can practice speaking with AI-powered avatars and play a variety of different games to review vocabulary (Figure 3). All environments in Immerse are highly interactive, meaning that learners can physically grab and manipulate different objects while collaborating with their peers. For instance, when learners practice ordering food in the fast-food restaurant scenario, they can ring orders up on a cash register, cook burgers on a grill, add condiments, fill up drinks, and even pay for orders by cash or card. Such a simulation is authentic and adequately reflects real-life scenarios, adding realism to the language learning environment.

One of the aims of this pilot study was to determine what types of activities worked most optimally in high school settings. Three different types of activities were tested with learners. The first activity was a live, teacher-led lesson in a fast-food restaurant where learners learned how to order food in French. The second activity was a peer-to-peer scavenger hunt in The Commons (i.e. Immerse's social lounge). The third activity was a structured, peer-to-peer activity in the shopping center where learners had to find and interact with certain objects while shopping together.

[1] https://www.oculus.com/experiences/quest/

[2] www.immerse.com

Figure 1. Learners interacting in a fast-food restaurant in Immerse.

Figure 2. Classroom view of learners interacting in VR.

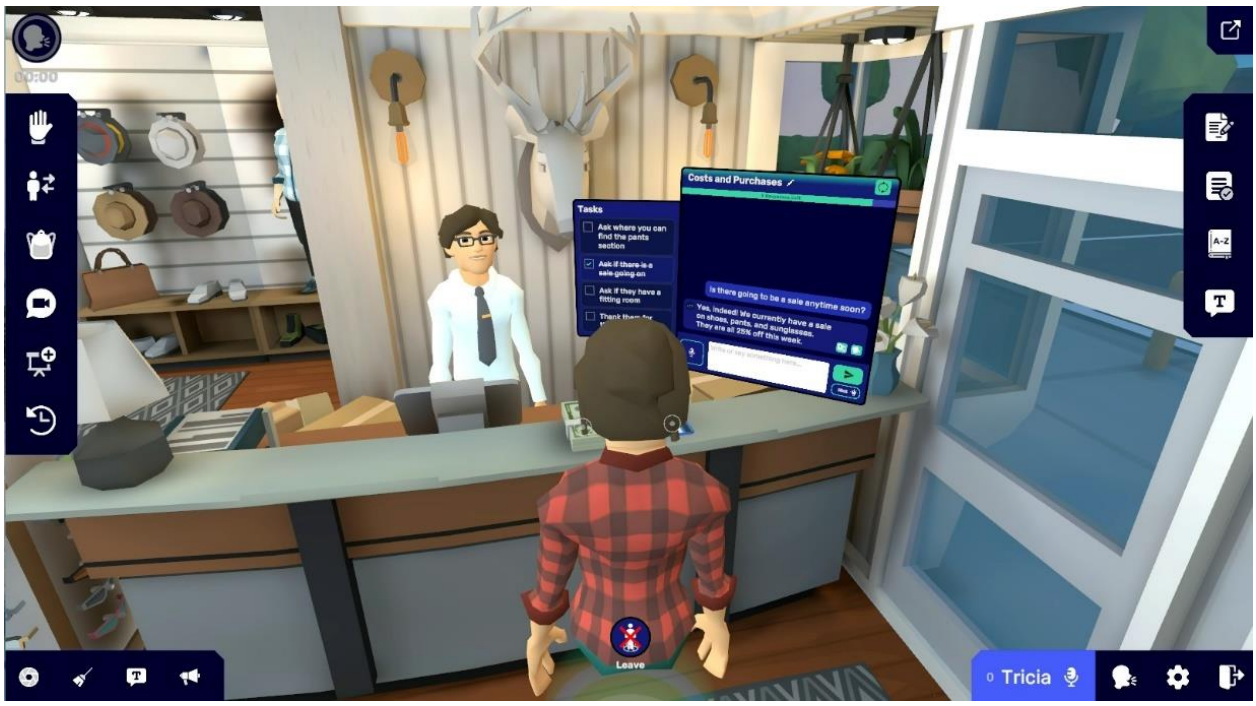


Figure 3. AI-powered avatar practice.

Following each activity, learners immediately completed a survey in which they were asked to describe their experience as "Awesome," "OK," or "Not so good." This survey was kept simple to allow learners to answer quickly and to not detract from class time. In addition, learners were asked to report on what they liked and disliked about the session (if anything). Learners' qualitative feedback was analyzed using MAXQDA software to identify trends in both positive and negative feedback.

3. Findings and discussion

Overall, 56.6% learners found VR "OK", 40.9% expressed thinking that it was "awesome", and only 3.3% perceived it as "not so good". Table 1 presents results based on school and activity (see Table 1).

Table 1. Learners' ratings of VR (all numbers are representative of percentages out of 100).

Rating	School		Scene/Activity			All Responses (n=123)
	Texas (n_1=109)	**Illinois** (n_2=14)	**Fast food Restaurant** (Teacher-led) (n_3=68)	**The Commons** (Learner-led) (n_4=14)	**Shopping Center** (Learner-led) (n_5=41)	
"Awesome"	36.1	78.6	33.8	78.6	39	40.9
"OK"	60.1	21.4	61.8	21.4	60	56.6
"Not so good"	3.7	0	4.4	0	1	3.3

Learners made 155 comments about what they liked and disliked. Of these, 103 were positive about VR and the Immerse application and 52 were negative. Table 2 presents a summary of the themes based on positive and negative feedback.

Table 2. Themes based on positive and negative feedback.

Positive Feedback	Negative Feedback
Interactivity with objects	Navigation
Interactivity with peers	Set-up time
Enjoyment	Physical discomfort
Visual appeal and realism	Audio feedback
Immersion	Distraction
Ease-of-use	
Break from traditional class	

With regard to positive feedback, learners commonly mentioned that they enjoyed being able to pick up, scan (to hear the pronunciation), and interact with various objects throughout scenes in Immerse (27%; 28 out of 103 comments). Along with object interaction, learners highlighted how they enjoyed being able to interact with their peers in VR together instead of being inside an isolated VR experience (21.3%). They also frequently said that VR lessons were a "fun" way to learn (19.4%) and expand their French vocabulary (13.6%). Last, learners mentioned that Immerse was visually impressive and realistic (7.8%), highly immersive (6.8%), and easy to use (4.9%). A few learners also pointed out that Immerse offered a nice break from their routine classroom activities (3.9%).

With regard to negative feedback, we identified five main concerns. First, the most common complaint (32.7%; 17 out of 52 comments) among learners was that VR and Immerse were difficult to navigate. For instance, some learners found the controls to be complicated. They also found it difficult to maneuver throughout the platform, which negatively impacted the set-up time. Second, 26.9% of learners mentioned the physical discomfort that can accompany using VR (e.g. tired eyes, nausea, headaches) and how this discomfort hindered their learning experience. Third, because learners were simultaneously inside Immerse and physically together in the classroom, acoustic feedback occurred (19.2%), leading to frustration and impeding their ability to hear others and effectively follow the lesson. Such issues, and the occasional lack of focus among their peers (9.6%), lowered the quality of the learning experience (17.3%).

Based on the lesson learned from the negative feedback we received from learners, we offer several recommendations for how to avoid possible issues when implementing VR on a larger scale in the classroom. Table 3 presents learners' main critiques and our suggestions for addressing them.

Table 3. Learners' critiques of VR and possible solutions.

Critiques	Solutions
Navigating VR and using controllers is difficult.	Provide learners with a more robust VR tutorial and additional practice sessions to reinforce the use of hand controllers. We recommend a minimum of two, one-hour sessions which should include a tutorial and hands-on practice.
Set-up time took away from class time in VR.	Provide both teachers and learners with more intense VR training so that they can easily start activities. Confirm that all headsets are properly connected to Wi-Fi and fully charged before class time.
VR headsets can cause physical discomfort.	Ensure that the learners select the best viewing setting in their VR headsets and that the width of the headset lenses is properly adjusted to their eyes. Make sure that learners have their motion settings set to 'Teleport'. Allow learners to decide whether they want to join VR sessions via a headset or a computer when possible.
Acoustic feedback was distracting.	Provide all learners with headphones to prevent audio feedback.
Classmates were not always on task during group activities.	Adapt post-activity questionnaires to include peer-assessment for participation and improve accountability during tasks.

4. Conclusions

The results of this pilot study of L2 French learners (n=123), who were high school students using the high-immersion VR app Immerse to learn French, revealed that after three exposures using the app, most learners had positive perceptions of their French learning experience (40.9% thought it was "awesome" and 56.6% felt it was "OK"). In response to the open-ended question about what they liked and disliked about the app, 66% of the comments were positive and 34% were negative. The positive features included the unique affordances of VR such as the ability to interact with both objects and peers and the realism and immersivity of the environment, whereas the criticisms provided helpful lessons learned for both teachers who want to use VR in their classrooms and researchers who study the effectiveness of VR apps. Key recommendations include the need to provide both learners and teachers with multiple and thorough training sessions, explicitly instructing users on the best viewing and motion settings while in VR, and carefully designing learning activities accompanied with questionnaires to assess the effectiveness of the activities.

Acknowledgements

We would like to thank Meta and Immerse for their support and funding of this project.

References

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of virtual reality language learning. *TechTrends, 3*. https://doi.org/10.1007/s11528-022-00717-w

Dooly, M., Thrasher, T., & Sadler, R. (2023). "Whoa! Incredible!:" Language Learning Experiences in Virtual Reality. *RELC Journal*. https://doi.org/10.1177/00336882231167610

Gruber, A., Canto, S., & Jauregi Ondarra, K. (2023). Exploring the use of social virtual reality for virtual exchange. *ReCALL*, 1-16. https://doi.org/10.1017/S0958344023000125

Kaplan-Rakowski, R., & Gruber, A. (2023). An experimental study on reading in high-immersion virtual reality. *British Journal of Educational Technology*, https://doi.org/10.1111/bjet.13392

Kaplan-Rakowski, R. (2011). Teaching foreign languages in a virtual world: Lesson plans. In G. Vincenti & J. Braman (Eds.), *Multi-user virtual environments for the classroom: practical approaches to teaching in virtual worlds* (pp. 438-453). IGI Global.

Lan, Y. J. (2020). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1), 1–15. http://hdl.handle.net/10125/44704

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(3), 947–958. https://dx.doi.org/10.1007/s10648-020-09586-2

Papin, K., & Kaplan-Rakowski, R. (2022). A study on vocabulary learning using immersive 360° pictures. *Computer Assisted Language Learning*, *35*. https://doi.org/10.1080/09588221.2022.2068613

Sadler, R. W., & Dooly, M. (2012). Language Learning in Virtual Worlds: Research and Practice. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary Computer-Assisted Language Learning* (Contemporary Studies in Linguistics). Bloomsbury Academic.

Tai, T. Y., & Chen, H. H. J. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research*, *59*(7), 1272–1293. https://doi.org/10.1177/0735633121994291

Taguchi, N. (2022). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, *56*(1), 308–335. https://doi.org/10.1002/tesq.3070

Thrasher, T. (2022). The impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility: An exploratory study. *CALICO Journal*, 39(2), 219–238. https://doi.org/10.1558/cj.42198

Ye, Y., & Kaplan-Rakowski, R. (2024). Practicing listening comprehension skills in high-immersion virtual reality. *SSRN*. https://ssrn.com/abstract=4335690

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16997

Writing with automatic speech recognition: Examining user's behaviours and text quality (lexical diversity)

Walcir Cardoso[a] and Danial Mehdipour-Kolour[b]
[a]Department of Education Concordia University, walcir.cardoso@concordia.ca and [b]Department of Education, Concordia University, danial.mehdipourkolour@mail.concordia.ca

How to cite: Cardoso, W.; Mehdipour-Kolour, D. (2023). Writing with automatic speech recognition: Examining user's behaviours and text quality (lexical diversity). In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16997

Abstract

This study explores the potential of Automatic Speech Recognition (ASR) as a writing tool by investigating user behaviours (strategies henceforth) and text quality (lexical diversity) when users engage with the technology. Thirty English second language writers dictated texts into an ASR system (Google Voice Typing) while also using optional additional input devices, such as keyboards and mice. Analysis of video recordings and field observations revealed four strategies employed by users to produce texts: use of ASR exclusively, ASR in tandem with keyboarding, ASR followed by keyboarding, and ASR followed by both keyboarding and ASR. These strategies reflected cognitive differences and text generation challenges. Text quality was operationalized through lexical diversity metrics. Results showed that ASR use in tandem with keyboarding and ASR followed by both keyboarding and ASR yielded greater lexical diversity, whereas the use of ASR exclusively or ASR followed by keyboarding had lower diversity. Findings suggest that the integrated use of ASR and keyboarding activates dual channels, thus dispersing cognitive load and possibly improving text quality (i.e. lexical diversity). This exploratory study demonstrates potential for ASR as a complementary writing tool and lays groundwork for further research on the strategic integration of ASR and keyboarding to improve the quality of written texts.

Keywords: *Automatic Speech Recognition (ASR), writing, user's behaviour, writing strategies, lexical diversity.*

1. Introduction

This study explores a breakthrough in the development of writing: the use of one's voice, made possible via Automatic Speech Recognition (ASR). ASR is an application found in most modern devices (e.g. computers, smartphones) as well as in word-processing and communication software, such as Google Docs. ASR converts speech into text in real-time: via a microphone, the software receives and identifies the words a person speaks, analyzes it using a set of algorithms, and finally produces an output in the form of a text. In a survey conducted by Enge (2020), he found that 45% of U.S. respondents use their voices to input text and for commands daily, indicating that people are becoming increasingly accustomed to speaking to their devices, particularly after the rise of voice-activated virtual assistants, such as Amazon Alexa and Apple Siri.

Despite the growth in ASR use for writing, research in this field has not received the attention it deserves. While studies indicate that ASR is a useful pedagogical tool for improving the pronunciation of second/foreign language

(L2) learners (e.g. Cardoso, 2022; McCrocklin, 2018; Mehdipour-Kolour & Cardoso, in press), very few studies have fully explored the use of ASR as a writing tool. Notable exceptions include Haug and Klein (2018), who compared handwritten and ASR-created argumentative texts among elementary school students, and Johnson and Cardoso (in press) and Mehdipour-Kolour and Cardoso (2023), who examined users' perceptions of ASR as a writing method. Indeed, most of the research available on ASR for writing involves users with learning difficulties (e.g. Ballard et al., 2019; Le et al., 2018; Quinlan, 2004).

As research on using ASR for composing texts is still emerging, this study aims to build foundational knowledge by examining the technology's feasibility as a writing tool (for the rationale, see Cardoso, 2022). Rather than focusing on user perceptions, which are explored in recent related work (Johnson & Cardoso, in press; Mehdipour-Kolour & Cardoso, 2023), this research investigates the core capabilities of an accessible ASR system (Google Voice Typing) by analyzing: (1) writers' behaviours (a component of *learning strategy*, defined as the behaviors or steps that learners use to improve their learning; Oxford, 1990), and (2) the correlations of behaviour with text quality. In this study, text quality is operationalized via lexical diversity (Yu, 2010; Kyle et al., 2021), a measure of how many different words are used and the relative frequency of each word. This measure is regarded as an important indicator of writing knowledge (Yu, 2010) and text quality (Laufer & Nation, 1995).

This study explores the potential of ASR as a writing tool by examining the behaviours of users while engaged in ASR-based writing and the quality of the texts produced (operationalized as *lexical diversity*). The study was guided by the following Research Questions (RQs):

1. How do writers interact with the ASR to compose texts? Specifically, what strategies do they use to compose texts (e.g. exclusive use of ASR, combining keyboarding with ASR)?
2. Is there a connection between the strategy selected and the quality of the texts produced (i.e. in terms of lexical diversity)?

2. Method

A total of 30 adult participants were recruited to participate in this study. They were English L2 speakers and had a variety of first languages (e.g. Chinese, Farsi, French, Portuguese) and had no previous experience using ASR for composing narrative texts.

The experiment took place remotely using Zoom and lasted approximately two hours. Participants first received an introduction to Google Voice Typing (GVT)(accessed by navigating to *Tools* and then *Voice Typing* in Google Docs) and its writing capabilities, including verbal cues such as *comma* and *period* for punctuation, and *new paragraph* and *underline last word* for formatting text. They were then instructed to compose texts by dictating into GVT while using any additional resources available to them, such as a keyboard, mouse, or handwriting. There were no restrictions on strategies or tools that participants could leverage to plan, monitor recognition, correct errors, and revise content. This enabled the observation of natural behaviors when using ASR for composing texts. The collected texts had an average length of about 300 words, with the mean of approximately 105.47 unique words across the 30 texts. Following insights from Holdsworth (2021) for a holistic analysis of users' experience and behaviours and thick descriptions (Geertz, 2008), data collection was triangulated to include: observation, fieldnotes, conversations with the participants, audio-visual materials (video recordings and screen capture of interactions with ASR), and an analysis of the texts produced to analyze lexical density.

To address the goals of the study, participants were given three narrative writing prompts (e.g. *First Love: Write a story about how you met your first love ...* – simplified due to space limitations) and asked to choose two to compose their narratives. At the end of the experiment, all texts were compiled for lexical diversity analysis. For this analysis, texts were imported to the Tool for the Automatic Analysis of Lexical Diversity (TAALED) (Kyle et al., 2021), an analysis tool developed to compute a wide range of lexical diversity indices. Four indices were selected (for the rationale and details, see Kyle et al., 2021): Root Type-Token Ratio, Moving-Average TTR, Measure of Textual Lexical Diversity, and MTLD moving-average-wrap.

3. Results

Adopting an exploratory approach**,** this study analyzed video recordings and field notes to answer the question of how writers interact with ASR to produce their texts regarding strategies used (RQ1). As illustrated in Table 1, participants employed a variety of strategies, broadly categorized into four types: the use of (1) ASR exclusively (ASR-Only; *n*=2; or 7% of all participants), (2) ASR in tandem with keyboarding (ASR=KB; *n*=13; 40%), (3) ASR followed by keyboarding (ASR>KB; *n*=5; 17%), and (4) ASR followed by both keyboarding and ASR (ASR>ASR=KB; *n*=10; 33%).

To answer RQ2 on whether there is a correlation between the ASR writers' choice of strategies and the lexical diversity of their written text, Eta and Eta squared tests were conducted. These tests are suited for analyzing categorical data and measuring effect size, and effective for exploring the correlation between strategy use and lexical diversity. We calculated the lexical diversity of participants' texts using TAALED (Kyle et al., 2021). Five levels of lexical diversity were identified in the collected narratives. The levels 'below average,' 'slightly below average,' 'average,' 'slightly above average,' and 'above average' were assigned based on predefined scores ranging from 'below average' (<50) to 'above average' (>50), as shown in Table 1.

Table 1. Writing strategies and associations with lexical diversity

			Lexical Diversity in Texts				
			Below	Slightly below	Average	Slightly above	Above
Writing Strategies	ASR-Only	*n*=2	2	0	0	0	0
	ASR=KB	*n*=13	1	1	0	7	4
	ASR>KB	*n*=5	3	1	0	1	0
	ASR>ASR=KB	*n*=10	0	1	1	4	4
Total		*n*=30	6	3	1	12	8

The findings of the Eta and Eta-squared tests are illustrated in Table 2, indicating that using ASR-Only corresponds to a significant moderate negative effect on lexical diversity, while ASR=KB exhibits a moderate positive association, but it lacks statistical significance in its effect. The adoption of ASR>KB is significantly and strongly correlated with a low level of lexical diversity. Finally, the use of ASR>ASR=KB demonstrates a moderate positive relationship with high lexical diversity, albeit without a statistically significant effect.

Table 2. Summary of the measures of associations and their effect sizes

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

4. Discussion and conclusions

This study examined L2 writers' interactions with ASR during text composition, specifically focusing on the strategies employed (RQ1), and explored whether a connection exists between the chosen strategy and the resultant quality of the produced texts (RQ2).

The findings for RQ1 indicate that L2 writers use a variety of strategies when composing with ASR technology. Four distinct approaches were identified through an analysis of participant behaviours: (1) the use of ASR exclusively, (2) ASR in tandem with keyboarding, (3) ASR followed by keyboarding, and (4) ASR followed by both keyboarding and ASR. The diverse array of strategies observed implies that writers exhibit distinct writing and, potentially, learning preferences. These preferences can be attributed to Individual Differences (IDs), which encompass a spectrum of factors such as varying cognitive aptitudes and learning styles (e.g., visual, auditory). Dörnyei (2005) defines as the distinctive traits and attributes that set individuals apart from one another. For example, considering the participants who used ASR exclusively for composing their texts, it is possible to speculate that these participants have heightened auditory and oral abilities in comparison with their peers. This ability led them to rely on ASR throughout all phases of the writing process, including planning, drafting, revising, and editing (for a discussion of these types of learners in SLA, see Pawlak & Kruk, 2022).

Regarding RQ2, Eta (η) and Eta-squared (η2) tests were conducted to investigate the connection between the writing strategies employed and the lexical diversity of the collected texts. The findings indicate that the use of ASR *combined* with keyboarding, namely ASR=KB and ASR>ASR=KB, resulted in a higher level of lexical diversity, as compared to the two other strategies. The latter strategies, which did not combine ASR and keyboarding, were significantly correlated with lower levels of lexical diversity as a measure of text quality. This finding confirms the Modality Effect Theory proposed by Leahy and Sweller (2011), who hypothesize that using several input modalities (e.g. visual and auditory channels) might assist in evenly distributing the cognitive burden, consequently yielding enhanced writing performance. Incorporating these strategies, the participants could cultivate ideas and reduce the need for constant text monitoring and repetitive corrections compared to exclusively using ASR, resulting in texts with higher lexical diversity. Considering that the results favouring ASR=KB and ASR>ASR=KB as writing strategies were not statistically significant due to the small sample size, cautious generalizations can still be made. One such generalization is that our findings highlight the potential benefits of leveraging ASR as a *complementary* rather than a standalone tool in the writing process.

As an exploratory study that aimed to provide initial insights into L2 writer strategies and lexical diversity patterns in ASR-based writing, there are certain limitations that need be acknowledged. These include the small sample size and the short duration of the study. Future research should examine these effects over an extended period of time, with more participants, and using additional textual measures beyond lexical diversity. However, the findings lay important groundwork for understanding how students strategically use ASR when writing, and how to potentially incorporate ASR effectively alongside typing for enriching text production with lexical variety. This has relevant implications for writing instruction and the use of ASR in L2 education.

Acknowledgements

We sincerely thank all the participants who generously contributed their time and insights to make this research possible. We also acknowledge partial funding from the *Social Sciences and Humanities Research Council of Canada* (Grant #430-2022-00512).

References

Ballard, K., Etter, N. M., Shen, S., Monroe, P., & Tan, C. T. (2019). Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for Apraxia of Speech Plus Aphasia. *American Journal of Speech-Language Pathology, 28*, 818–834. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0109

Cardoso, W. (2022). Technology for speaking development. In T.M. Derwing., M.J. Murray., & R.I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 299–313). Routledge. https://doi.org/10.4324/9781003022497

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

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

Geertz, C. (2008). Thick description: Toward an interpretive theory of culture. In T. Oakes, & P. L. Price (Eds.), *The cultural geography reader* (pp. 41–51). Routledge.

Haug, K.N., & Klein, P.D. (2018). The effect of speech-to-text technology on learning a writing strategy. *Reading & Writing Quarterly, 34*(1), 47–62. https://doi.org/10.1080/10573569.2017.1326014

Holdsworth, J. (2021). *Doing ethnographic research: a practical guide*. Sage Publications.

Johnson, C. & Cardoso, W. (in press). Hey Google, let's write: Examining L2 learners' acceptance of automatic speech recognition as a writing tool. *CALICO Journal*.

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

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

Le, D., Licata, K., & Provost, E.M. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication, 100*, 1–12. https://doi.org/10.1016/j.specom.2018.04.001

Mehdipour-Kolour, D. & Cardoso, W. (in press). A systematic literature review of automatic speech recognition in L2 learning: A case for L2 writing. In M. Peterson & N. Jabbari (Eds.), *Routledge Frontiers in computer assisted language learning* (17 pages). London, UK: Routledge. https://doi.org/10.4324/9781003395218

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers. https://escholarship.org/uc/item/1446j36q

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

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

McCrocklin, S. (2018). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal, 36*(2), 119–137. https://doi.org/10.1558/cj.34738

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

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16958

Assessing 21st century digital literacies in Japanese higher education

Sandra Healy[a] and Olivia Kennedy[b]
[a]Kyoto Institute of Technology, Kyoto, , healy@kit.ac.jp and [b]Nagahama Institute of Bioscience and Technology, Nagahama, , Shiga, o_kennedy@nagahama-i-bio.ac.jp

How to cite: Healy, S.; Kennedy, O. (2023). Assessing 21st century digital literacies in Japanese higher education. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EUROCALL2023.2023.16958

Abstract

This study examines how the seven digital learner profiles included in the International Society for Technology in Education (ISTE) standards for students apply to 193 participants in Japan, utilizing a questionnaire developed by Mills and White (2023). These profiles explore a broad range of 21st century digital skills, competencies, and literacies essential for learners to be successful in this digital age, and include both creativity and communication. The results revealed positive identification with the learner profiles (M=3.874). While Japanese learners are often found to lack digital skills, the highest mean values were associated with the profiles of Digital Citizen, Knowledge Constructor, and Computational Thinker. In contrast, the profiles of Empowered Learner, Innovative Designer, Creative Communicator, and Global Collaborator demonstrated lower mean values, particularly in more creative and communicative aspects. To ensure students are well-prepared for the digital era, therefore, fostering creativity and communication skills is crucial. There have been a lack of studies that measure the various aspects of 21st century digital skills (van Laar et al., 2020), and this study aims to add to the literature.

Keywords: *digital literacies, 21st century digital skills, creativity, communication skills.*

1. Introduction

In the rapidly evolving landscape of the 21st century, digital literacies initially focused on technical skills now encompass a wide range of competencies, including creativity and communication. Educational policymakers have initiated reforms to foster creative thinking and incorporate these skills into education systems; however, this endeavour has posed persistent challenges worldwide (van Laar et al., 2020), including in the Japanese context.

Despite the Japanese government's efforts to reform English language education, most still learn English through the grammar-translation method. Characterized as highly teacher-centred and examination-focused (Aubrey, 2020), this method not only fails to promote learner autonomy, active participation, and communicative abilities both in Japanese and in English, but also contributes to passive learning habits (Egitim, 2022).

Japan has also been slow to integrate 21st century digital skills (Funamori, 2017; Terashima, 2019) into educational policy, and so learner skills are similarly underdeveloped (Mills & White, 2023). How to foster learner creativity has also been an ongoing issue, and the absence of a clear definition of the concept, combined with the existing policy framework, has limited the effective promotion of creative skills among students (Smith, 2018).

The International Society for Technology in Education (ISTE) standards for students focus on leveraging technology in education to create impactful, sustainable, scalable, and equitable learning experiences for all

learners. They specifically aim to empower students by giving them a voice in their learning process, ensuring that learning becomes a student-driven endeavour. There are seven ideal learner profiles:

1. An *Empowered Learner* leverages technology to take an active role in choosing, achieving, and demonstrating competency in their learning goals.

2. A *Digital Citizen* recognizes the rights, responsibilities, and opportunities of living in an interconnected digital world, and acts in safe, legal, and ethical ways.

3. A *Knowledge Constructor* critically curates resources using digital tools to construct knowledge, produce artefacts, and create meaningful learning experiences.

4. An *Innovative Designer* uses technologies to identify and solve problems by creating new, practical, imaginative solutions.

5. A *Computational Thinker* develops and employs strategies for understanding and solving problems using technology.

6. A *Creative Communicator* communicates clearly and creatively using appropriate platforms, tools, styles, formats, and digital media.

7. A *Global Collaborator* uses digital tools to broaden their perspectives and enrich their learning by collaborating with others locally and globally.

This study explores 21st century digital literacies with a specific focus on the significance of creativity and communication in the context of the Japanese education system, and asks the question: To what extent do Japanese university students identify with the learner profiles outlined in the ISTE Standards for students, particularly in terms of creativity and communication?

2. Method

The 193 Japanese participants in this study are from a national university in Japan that focuses on Science, Technology, Engineering, Arts and Mathematics (STEAM) education. Most of the participants were male, 83%, aged 18 or 19 years old and in their first year at university.

The survey consists of 24 items from a Likert scale instrument developed by Mills and White (2023) with responses ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) based on the seven learner profiles from the 2017 ISTE Standards (ISTE, 2017).

Data collection occurred in July 2022 using a Google Forms document on the university Learning Management System. The surveys were provided in both English and Japanese. Participants gave consent for their anonymised data to be used for research purposes. The resulting data was entered into Microsoft Excel and then transferred to JASP (JASP, 2023) for analysis.

3. Results

All the scales and subscales were acceptable with a Cronbach's alpha of .879, higher than the recommended 0.70 (George & Mallery, 2003). The mean and standard deviations are shown in Table 1 for each learner profile.

Table 1. Mean and standard deviation of all questions

Questions	Mean	SD	Learner profiles
1. I use technology to define and achieve my educational goals.	3.743	1.245	Empowered learner

2. I use technology to seek feedback that will improve my learning.	4.029	0.985	Empowered learner
3. I use technology to demonstrate my learning in a variety of ways.	3.743	1.197	Empowered learner
4. I seek out new technologies to achieve my educational goals.	3.8	1.132	Empowered learner
5. I use the Internet and social networking sites in a positive, legal, and ethical way.	4.2	0.833	Digital citizen
6. I manage my digital identity and I am aware of the permanence of my actions in the digital world.	4.571	0.698	Digital citizen
7. I understand the importance of keeping my personal data safe to maintain digital privacy and security.	4.657	0.684	Digital citizen
8. I use technology to search for information that contributes to my intellectual and creative growth.	4.257	0.657	Knowledge constructor
9. I build knowledge by actively exploring real-world issues and problems, developing ideas and theories, and pursuing answers and solutions.	4.029	0.857	Knowledge constructor
10. I evaluate the accuracy, perspective, credibility, and relevance of information, media, data, or other resources.	4.057	0.639	Knowledge constructor
11. I use technology to generate new ideas, test theories, develop works of art, and solve real problems.	3.857	1.167	Innovative designer
12. I have a tolerance for ambiguity, and have the capacity to work on open-ended problems.	3.429	0.815	Innovative designer
13. When creating something new I continually test, reflect, and improve the product until I am satisfied with the result.	3.8	0.759	Innovative designer
14. I use technology to solve practical problems through a logical process.	4.314	0.758	Computational thinker
15. I can collect and analyze data using digital tools and present data in a way that demonstrates my problem-solving and decision making.	4.114	0.583	Computational thinker
16. I divide problems into their parts, extract key ideas, and develop descriptive models to understand systems or assist in problem-solving.	3.629	0.808	Computational thinker
17. I use a variety of technologies, visual instruments, models, and simulations to communicate complex ideas to others clearly and effectively.	3.629	1.114	Creative communicator
18. I create original works or responsibly repurpose or remix digital resources into new projects.	3.514	1.314	Creative communicator
19. I explain complex ideas clearly and effectively by creating or using a variety of digital solutions.	3.429	1.145	Creative communicator
20. I am able to modify my message or content for a particular audience.	3.971	0.822	Creative communicator

21. I use digital tools to communicate with learners from different backgrounds and cultures to exchange experiences and understand them.	3.571	1.065	Global collaborator
22. I use technology to work with others (peers, teachers, and others) to discuss problems from various viewpoints.	4.029	0.785	Global collaborator
23. I contribute constructively to team projects by assuming a role and responsibility to work towards a common goal.	3.371	1.239	Global collaborator
24. I explore local and global issues using technologies to work with others.	3.229	1.114	Global collaborator

4. Discussion

While previous research has found Japanese learners to be largely passive (Egitim, 2022), overall responses to the survey were positive, showing that the participants perceived themselves as empowered 21st century learners able to use technology to achieve their goals. As such, all items in *Empowered Learner* category yielded high scores, with the highest mean observed for Question 2 (using technology to seek feedback for learning improvement) showing not only learners' proactive learning strategies, but their awareness of them. It also reflects the concept of *kaizen*, continual improvement, in Japanese culture (Khayum, 2015). This finding is of specific interest in the CALL field: learners confident in the use of technology are well-placed to access the wealth of CALL resources now available to achieve language learning goals.

The participants' understanding of the importance of using technology safely and ethically, consistent with Japanese norms emphasizing social and collective responsibility (Yamamoto & Lloyd, 2019) seen in the high scores yielded by all items in the *Digital Citizen* category, is also of note. The third category with consistently high mean scores was *Knowledge Constructor*. Question 8, which focuses on using technology to search for information that contributes to intellectual and creative growth, received the highest score in this category. This is in line with the usage of technology in Japanese education for research rather than creation and collaboration (Mougenot, 2016).

The categories with the lowest mean values were related to communication, collaboration, and creativity. Whilst Japan has made efforts to improve all three in higher education (Nyugen, 2019), it appears from these results that there needs to be an increased focus on these areas. For example, while participants exhibited positivity in the *Innovative Designer* category, especially in terms of generating ideas, testing theories, and solving problems using technology, all can be achieved individually rather than in collaboration with peers. In connection with this, participants also expressed confidence in using technology to solve practical problems in the *Computational Thinker* category, but they displayed uncertainty when it came to breaking down complex problems and developing models to aid problem-solving. This can also be seen in the results for Question 18 in which the participants expressed less confidence in creating original works, again pointing to difficulties surrounding active creation. One aspect related to this is the Japanese preference for avoiding uncertainty (Yamamoto & Lloyd, 2019). This risk aversion is further evident in the lowest mean value observed for Question 12, which evaluates learners' ability to tolerate ambiguity and work on open-ended problems. On the other hand, participants felt capable of modifying their messages or content for specific audiences (Question 20, *Creative Communicator* category), possibly reflecting the influence of Japanese linguaculture in which status is clearly reflected in language and so constant modification and adaption is necessary (Maruki, 2022).

Finally, the lowest mean score recorded in the *Global Collaborator* category was for Question 24, showing participants' lack of confidence in exploring local and global issues using technology. Question 22, however, about using technology to collaborate and discuss problems from different perspectives, yielded a comparatively high mean. The presence of the word 'teachers' in the question might have influenced participants by contributing to their sense of comfort. This could be due to a cultural inclination towards teacher dependence in Japanese culture.

Similarly, because Japanese students are taught to be more comfortable with discussion and research rather than action (Nyugen, 2019), the word 'discuss' may have also influenced participants' responses to this item.

Despite the valuable insights gained from this study, there are certain limitations. While efficient for gathering data, self-reporting questionnaires may suffer from potential biases, such as overestimation of abilities. Future research should therefore explore alternative methods to complement and validate these findings (Hargittai, 2005). Additionally, while this study examines how university students perceive their own digital literacy, studies that more specifically explore the interplay of creativity and communication in digital literacy need to be undertaken so that effective policies to achieve them can be developed. This article is part of one such larger project that aims to do so.

5. Conclusions

In contrast to existing research about Japanese students' learning styles, this study found that participants positively perceived themselves as independent and digitally literate 21st century learners. Even though the university where this research was undertaken is noted for its design-based, innovative curriculum (Mougenot, 2016), the balance between uncertainty, avoidance, and risk-taking in creative and online work needs to be further improved as the participants felt less confident in their communication, collaborative, and creative skills. To adequately prepare students for success in the technology-driven world, policymakers and educators must focus on developing these CALL skills alongside technical expertise. By nurturing such skills, educational systems can empower learners to become innovative problem solvers, effective communicators, and global collaborators, well-equipped to thrive in the dynamic digital era.

References

Aubrey, S. (2020). The role of task-based interaction in perceived language learning in a Japanese EFL classroom. In C. Lambert & R. Oliver (Eds.), *Using tasks in second language teaching: practice in diverse context* (pp. 281-305). Multilingual Matters.

Egitim, S. (2022). *Collaborative leadership through leaderful classroom practices: Everybody is a leader*. Candlin & Mynard.

Funamori, M. (2017). The issues Japanese higher education face in the Digital Age: Are Japanese universities to blame for the slow progress towards an information-based society. *International Journal of Institutional Research and Management, 1*(1), 37-51.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon.

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, *23*(3), 371–379. https://doi.org/10.1177/0894439305275911

ISTE (2017). ISTE standards. International society for technology in education. https://www.iste.org/standards

JASP Team (2023). JASP (Version 0.17.3) [Computer software].

Khayum, H. M. O. (2015). Kaizen: Potentiality in utilization of human prospects to achieve continuous improvement in the quality of higher education. *International Journal of Multidisciplinary and Current Research*, *3*, 1223-1229.

Maruki, Y. (2022). Keigo to use and to be used: Reevaluation of keigo learning in Japanese language classes. *Journal of Japanese Language Education and Linguistics*, *6*(2), 142-153.

Mills, D. J., & White, J. (2023). The influence of digital gameplay on learner profiles in the Japanese university context. [Unpublished manuscript].

Mougenot, C. (2016). Japanese higher education in a global context: Making students more innovation-minded. *工学教育 [Engineering Education], 64*(5), 39-45.

Nyugen, D. M. H. (2019). Embedding core values of creativity and teamwork in higher education: lessons learned from Japanese universities. I*nternational Journal of Education and Research, 7*(6), 199-206.

Smith, C. A. (2018). "Creativity" in Japanese education policy. In P. Clements, A. Krause, & P. Bennett (Eds.), *Language teaching in a global age: Shaping the classroom, shaping the world.* Tokyo: JALT, 8-14.

Terashima, K. (2019). Professional development for Middle leader teachers: ICT integration in schools in Japan. In K. Graziano (Ed.), *Society for Information Technology & Teacher Education International Conference* (pp. 2543-2548). Las Vegas, NV, United States: Association for the Advancement of Computing in Education (AACE).

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: a systematic literature review. *SAGE Open*, *10*(1). https://doi.org/10.1177/2158244019900176

Yamamoto, K., & Lloyd, R. A. (2019). Ethical Considerations of Japanese Business Culture. *Journal of Business Diversity, 19*(2), 113-122.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16922

Attitudes of part-time Japanese university teachers on technology use after emergency remote teaching

Frances Shiobara[a], Kym Jolley[b] and Mark Donnellan[c]

[a]Department of Education, Kobe Shoin Women's University, , fshiobara@shoin.ac.jp; [b]Faculty of Cultural and Expression Studies, Baika Women's University, , k-jolley@baika.ac.jp and [c]Faculty of Informatics, Kindai University, , donnellan@kindai.ac.jp

How to cite: Shiobara, F.; Jolley K.; Donnellan, M. (2023). Attitudes of part-time Japanese university teachers on technology use after emergency remote teaching. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16922

Abstract

With the onset of the COVID-19 pandemic in 2020, many Japanese universities made the decision to stop on-campus classes. This generally meant teaching remotely, either synchronously or asynchronously. Online courses are usually planned in advance with a curriculum and materials designed for that mode of teaching. The situation in 2020 was different, most teachers had only a few weeks to adapt their teaching materials and methods for Emergency Remote Teaching (ERT). The aim of this research was to investigate the experiences of university teachers during and after the period of ERT and to investigate how we can learn from this experience to avoid what Vegas (2022) refers to as a 'missed opportunity'. This was achieved through an initial pilot survey and five semi-structured follow-up interviews conducted in 2023. The results indicated that despite some reporting high levels of initial anxiety, most teachers adapted, learning new skills and developing their technical expertise, which is now being used in face-to-face classes. As most teachers reported support in the period of ERT as having been uncoordinated – being provided by colleagues rather than through official workshops organized by institutions - this is an area that could be developed in the future.

Keywords: *Emergency Remote Teaching, Higher Education, Online Learning.*

1. Introduction

Technology has had an enormous impact on most aspects of our lives, but prior to COVID-19 it did not appear to have had a major impact on education. Vegas (2022) hypothesized that this might be explained by the theoretical framework presented by Cohen and Ball (1999), which argues that the reason educational reform is usually not effective is the failure to pay enough attention to interactions between the educators and learners. Although this framework was aimed at educational reform in general, it could more specifically be applied to educational technology. Vegas (2022) found that ed-tech interventions are most effective when they incorporate some of the advantages of ed-tech, for example the possibility to increase scale of instruction, enable personalized instruction, expand possibilities for practice, and increase possibilities for engagement. Carefully planned online courses are usually built to incorporate these features, unfortunately in the rush to switch to ERT these factors may initially not have been incorporated (Hodges et al., 2020). However, over the period of ERT, teachers were able to better

incorporate effective ed-tech in their teaching. Unfortunately, what Vegas (2022) found worrying is that in many instances, as schools reopened, educational systems seemed to return to traditional pre-COVID-19 styles of education, leading to a lost opportunity to learn from what had been experienced.

Learning Management Systems (LMS) allow schools to offer classes both asynchronously and synchronously. They can be used to share teaching materials, collect assignments, set tests, and for communicating with students (Dovrat, 2022). During ERT these systems were widely used in Japanese universities in a way they had not been before. Continuing the use of LMS, even in face-to-face classes, allows class materials to be easily accessed by students and updated by teachers. It also allows for easy tracking of submissions and grades by teachers and students (Eto, 2021). Oliveira et al. (2021) found positive and negative perspectives emerged during ERT. Namely, teachers were found to be more available to answer student questions; but teachers and students lacked face-to-face interaction and human contact within class time, largely due to students not turning on their cameras. Oliveira et al. (2021) also found that teachers and students experienced an improved balance between their academic and personal lives due to reduced commuting times and attending conferences online. However, they also reported that teachers admitted to 'burn out' and mental health issues, while students reported that they pretended to attend lectures when they were really doing something else.

By acknowledging the negative and positive aspects of ERT during COVID-19, it is hoped that this research will assist educators in incorporating best practices in online and face-to-face classes in the future, and avoid Vegas's (2022) concerns that schools and universities will revert to pre-pandemic styles of teaching without adopting all that has been learned.

2. Method

2.1. Research method

In 2022 an initial pilot survey (*N*=31) was carried out using Google Forms to understand the experiences of university instructors in Japan during the ERT period (Donnellan et al., 2022). It revealed that part-time lecturers seemed to need the most support because they dealt with multiple LMS, class types, and policies at different institutions. The researchers decided to conduct follow-up semi-structured interviews to further investigate the different challenges and circumstances faced by part-time teachers during ERT.

2.2. Participants

Five part-time teachers who responded to the survey and expressed willingness to be interviewed were selected for follow-up interviews. Semi-structured interviews were conducted and transcribed using Zoom, each one lasting approximately 30 minutes. The researchers then coded the data into themes. Informed consent was obtained from each interviewee at the time of the initial pilot survey and again at the time of the interviews. The results of the qualitative interview data were compared to the initial survey data and analyzed to give clear insights into the types of training and support that is necessary post-ERT.

3. Results

A variety of opinions were shared by the interviewees, but a few of the pertinent recurring themes, those of 'support', 'money', 'work-life balance', and 'skills acquired', are explored below.

3.1 Support

All teachers stated that support from administration was limited and when it was provided it was mostly in Japanese and was not specifically related to language teaching. Most of the teachers stated that Japanese language ability was a problem for them or colleagues. Some coordinators tried to overcome this by interpreting and translating communications into English. Furthermore, all of the teachers interviewed found the written manuals for LMS to either be inadequate or written in Japanese that was very difficult to understand for non-native Japanese speakers. For support, the majority of teachers turned to online tutorials. It may be due to this that most teachers were more positive about LMS such as Google Classroom, Moodle, and Blackboard rather than bespoke LMS made for specific universities with little online support in English. Furthermore, in most cases teachers stated that while their coordinators were not specialists they did try to answer questions, but they felt that the best support came from peers and informal groups formed on social media platforms for those navigating similar issues.

All of the interviewees stated that at some point they actually trained or supported other teachers. Three of the teachers reported being very confident with technology and were asked to run various workshops for other teachers. This was based on the fact that they had used the technology before, even if that knowledge was limited. Indeed, Interviewee 1 commented, "I had used Zoom about 3 times before and was the one doing the training for some people, because I guess 3 times is better than 0 times." This indicates the inadequate level of knowledge about online teaching at many universities in Japan prior to ERT. Interviewee 5, who was not confident at the beginning of ERT, helped other teachers who were even less confident, saying, "There were a couple of other teachers that were struggling. So, then I was trying to help that teacher [sic] separately." The implication is that support was very organic, coming from teachers seeing a need and stepping in to support each other. It appears to have been a very community-based approach with Interviewee 3 noting, "The thing I would say is that it was obvious that we were all learning together." Two of the teachers mentioned receiving and providing psychological support as well as practical advice, as mentioned by Interviewee 4, "I was called for … less teaching and more psychiatric help."

3.2 Money

Teachers also commented on the financial cost of ERT. Most teachers received no funding for Wi-Fi or to buy a computer, despite the fact that the majority of students at universities in Japan were given financial support to buy computers (Kyodo News, 2020). One teacher was given ¥100 per online lesson to pay for electricity and another teacher was given a small lump sum to pay for extra materials. This may have helped, but was insufficient for teachers who needed to buy new hardware or upgrade their internet connection. In addition, Interviewee 3 said that they did not use purchasable apps or software because they had no funds to pay for them, stating, "I avoided using almost all the apps I think … the fact that you don't have money to buy stuff is an issue."

3.3 Work-life balance

All of the interviewees mentioned positive aspects of ERT, in particular, an improvement in their work-life balance. Two of the interviewees had young children and spoke of the benefits of being able to be at home with their children. Additionally, commuting became unnecessary, allowing greater flexibility with schedules. One of the interviewees commented on how much they loved this, saying they had time to go to Starbucks and see their child off each morning as they started kindergarten.

3.4 Skills acquired

The teachers stated that not only had they gained skills and confidence in using technology, but students did too. This has greatly impacted how post-ERT face-to-face classes are conducted. Students are more computer literate, and university administrations are more accepting of technology in the classroom. Interviewee 3 stated that their university now has a Media Center where teachers can receive advice on how to implement pedagogical ideas. The

use of different LMS at different universities did not seem to cause major issues, which surprisingly contrasts with initial findings in the pilot survey. It is hypothesized that this contrast is due to the fact that those more confident or knowledgeable with technology use for classes stated their willingness to be interviewed on the initial survey. Indeed, all the teachers still use the LMS extensively for testing and assignments post-ERT. Two of them even stated that they have gone paperless, just as they were during the pandemic. Communicating with students has also become much easier. At the beginning of the pandemic many students were using email to communicate with their teacher, but can now use messaging systems that are part of university LMS. Surprisingly most of the teachers in this research spoke very positively about ERT, using words such as "I loved it" or "I loved the challenge." On the other hand, hybrid classes with students joining both online and in person were very unpopular, with Interviewee 1 stating, "Students could decide 1 minute before class which one they were doing … just awful, because I had no idea how to get groups sorted and stuff like that. I would never want to do hybrid again."

4. Discussion

Though it might be assumed that ERT was very stressful, teachers in this research were very positive. Comments about being able to spend more time with their children and not commute were common. Teachers also formed close bonds supporting each other. In addition, though it may have been expected that university administrations would have provided support for teachers, it appears that on the whole there was very little technological or financial support. Teachers joined together forming online groups to teach each other what they knew. Much of this seemed to be people with little knowledge guiding people with even less knowledge. This mutual support among peers was not formalized but, as such, seemed to work effectively. It may have been that university administrators were more focused on students and trying to manage the logistics of updating LMS and Wi-Fi networks, and therefore did not have the resources to devote to teacher training. Results from the initial pilot survey showed that using different LMS at multiple universities caused a problem for the part-time teachers, but they did not cite this as a major obstacle to successful ERT during the interviews. However, only one of the teachers interviewed stated that they were not confident with technology prior to ERT, this may be because teachers interested in technology were more likely to fill in the initial survey and subsequently agree to be interviewed.

5. Conclusions

Although this research seems to indicate that the period of ERT was successful, it is obvious that in the beginning there was a period when universities and part-time teachers were struggling. The fact ERT ended up being fruitful for those in this research seems to be largely due to individual teachers helping each other. In the future, it would be advantageous for universities to build a system for training all teachers and students on effective use of LMS and digital tools. Financial support for part-time teachers was also lacking. This is an ongoing issue as part-time teachers do not receive special funds that allow them to utilize tools they found effective during ERT or think could have a positive impact in their classes post-ERT, possibly leading to more missed opportunities. It is hoped that in the future, universities will find ways to support all teachers so that technology will complement classroom teaching rather than replace it.

Although this research has been informative, in future, it would be useful to investigate the experiences of teachers who were not confident with technology use to find out how they coped during the ERT period.

References

Cohen, D. K., & Ball, D. L. (1999). *Instruction, Capacity, and Improvement.* Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education. https://www.cpre.org/sites/default/files/researchreport/783_rr43.pdf

Donnellan, M., Shiobara, F., & Jolley, K. (2022). An Investigation of Emergency Remote Teaching Conditions and EFL at Japanese Universities. *Ikoma Journal of Economics, 20*(2), 83-111. http://doi.org/10.15100/00023299

Dovrat, L. (2022). Perceptions of emergency remote teaching tools used during Covid-19 online teaching by an Israeli English for Academic Purpose (EAP) department. *Studies in Technology Enhanced Learning, 2*(2), 273-287. https://doi.org/ 10.21428/8c225f6e.53d353b7

Eto, Y. (2021). Students' behaviors and perceptions using an LMS. *Accents Asia, 14*(1), 28-36. https://www.accentsasia.org/previous-issues/

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020, March 27). The Difference between emergency remote teaching and online learning. *EDUCAUSE Review*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Kyodo News. (2020, April 25). *Japan universities to help subsidize students' costs for online learning.* Kyodo News. https://english.kyodonews.net/news/2020/04/b281e206e86b-japan-universities-to-help-subsidize-students-costs-for-online-learning.html

Oliveira, G., Teixeira, J., Torres, A., & Morais, C. (2021). An exploratory study on the emergency remote education experience of higher education students and teachers during the COVID-19 pandemic. *British Journal of Educational Technology, 52*, 1357-1376. https://doi.org/10.1111/bjet.13112.

Vegas, E. (2022, March 11). *Education technology post-COVID-19: A missed opportunity?* Education Plus Development. https://www.brookings.edu/blog/education-plus-development/2022/03/11/education technology-post-covid-19-a-missed-opportunity/

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16990

Persian FLAIR: grammatically intelligent web search for language learning

Evan Bartholomeusz[a] and Robert Reynolds[b]
[a]College of Life Sciences, Brigham Young University, ejbart@byu.edu and [b]College of Humanities, Office of Digital Humanities, Brigham Young University, robert_reynolds@byu.edu

How to cite: Bartholomeusz, E.; Reynolds, R. (2023). Persian FLAIR: grammatically intelligent web search for language learning. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16990

Abstract

We describe our work on the development of a new Persian module for FLAIR, a Java-based grammatically intelligent web search engine (Chinkina et al., 2016; Chinkina & Meurers, 2016). This website allows Persian teachers and Persian language learners to search for texts about any desired topic, prioritizing documents that are rich in selected grammatical constructions, and filtering them based on text difficulty. Currently, the Persian module includes 56 unique grammatical constructions. Our Persian FLAIR module allows for a more self-guided approach to language learning and provides a level of flexibility to both language learners and teachers. Students are able to read accessible texts that interest them, while teachers are able to find real world examples of grammatical concepts used together in the same text. FLAIR can be used to create custom learning materials. Teachers can use FLAIR to search for texts containing particular grammar concepts being focused on in the classroom, which can then be used as relevant topics of discussion or reading. We developed FLAIR in a response to the limited availability of resources for learning and teaching Persian, particularly when it comes to its colloquial forms. With FLAIR, we hope to provide a valuable tool for students and teachers alike.

Keywords: *Persian, web search, authentic text, natural language processing.*

1. Introduction

Persian, also known by its endonyms of Farsi, Dari, and Tajik, is a less commonly taught language with over 130 million L1 and L2 speakers worldwide (Windfuhr, 2009). It is the official language of Iran, Afghanistan, and Tajikistan, and is also spoken in Uzbekistan, Turkmenistan, and Iraq. Persian continues to hold significance and relevance, not only through its political influence in the Middle East, but also because of the steadily growing Iranian and Afghan diasporas. Despite Persian's worldwide prevalence and importance, it remains a low-resource language. Many universities lack Persian language courses. Most available textbooks focus on formal, prescriptivist approaches and tend to ignore practical applications of colloquial Persian. Current CALL resources for Persian consist mainly of vocabulary apps, with limited resources for pronunciation, grammar, and relevant cultural topics.

Form-focused Linguistically Aware Information Retrieval (FLAIR) is an architecture created by Chinkina, Kannan, and Meurers (2016) to provide learners of English and German with easily accessible comprehensible input in line with the learner's interests. FLAIR uses Microsoft Bing API to fetch results for the user's desired

query, after which the web pages are parsed using Stanford CoreNLP's library (Manning et al., 2014). Research suggests that L2 acquisition occurs best when the learner is made consciously aware of language structure during L2 exposure (Schmidt, 1990). FLAIR allows teachers to find appropriate texts for instruction based on grammatical forms being discussed in the classroom. It also allows learners to independently explore the relationship between form and function in the language without the need for a traditional classroom setting. By developing a similar tool for Persian, we hope to improve the language learning experience for both teachers and learners and to make Persian more accessible to those who are interested in learning.

In this paper, we describe the purpose of FLAIR and specific changes made to adapt its architecture to Persian. Our module is free[1] and open-source[2].

2. Architecture

FLAIR provides personalizable, interactive input for language learners, which has been shown to facilitate self-correction in L2 production (Gass & Varonis, 1994). Ranking by text difficulty allows language learners to focus on texts best adapted to their current abilities and proficiency levels in the language. Research shows that language learners who are exposed to input at the appropriate difficulty level progress faster than those whose language input is far above or far below their actual language capabilities (Krashen, 1977; Swain, 1985).

FLAIR relies on a combination of machine learning and rule-based categorization to provide users with accurate, relevant search results. By searching for relevant grammatical forms (verb tenses, prepositions, pronoun types, etc.) including those not found in English, users can easily locate texts that showcase desired grammar concepts. This provides them with exposure to natural language constructions used in everyday contexts. FLAIR's architecture then categorizes the texts based on difficulty, allowing users to find level-appropriate learning material. FLAIR is designed to be highly localizable and adaptable to fit any language, and has already proven effective in other languages, including English,German (Chinkina et al., 2016), Russian (Reynolds, 2022), and Arabic (Hveem, 2019). A user selects a search language, enters a query, selects the desired number of results, then the search continues until the desired number of sites with a high enough text content have been found. FLAIR then converts the websites into documents that can be parsed by the selected natural language processing pipeline, which annotates every occurrence of target grammatical constructions.

Once the document is parsed, the user can adjust settings to rank the documents based the in-text frequency of selected grammatical features, as well as text difficulty based on the CEFR scale. Currently, our analysis relies on a modified Fleisch-Kincaid coefficient, known as Fleisch-Dayani (Dayani, 1990), to determine difficulty level. Users can select multiple grammatical features simultaneously, placing different weights on the ranking of different features. Ranking the texts by grammar concept allows the user to focus on both form and function in the language, observing the interplay between grammar and context in authentic and natural settings. For example, a Persian language teacher covering the Iranian parliament in class, and wanting to teach students about subjunctive verbs, can search for "مجلس ایران" (Iranian parliament), and prioritize texts containing frequent subjunctive verb usages. The teacher can then assign these texts to the students for take-home reading, asking them to highlight or pay attention to each instance of subjunctive verbs and then go online to check Persian FLAIR's highlighted list of all subjunctive verbs to check their work.

3. FLAIR for Persian

Compared to other languages from the Middle East, Persian has a relatively simple grammar. There is no grammatical gender, very few cases, and verb conjugation in all tenses is almost entirely regular. These features lend themselves well to the rule-based aspects of FLAIR's architecture. We created a list of useful grammar

[1] An instance of the server can be accessed at https://icall.byu.edu/flair-2.0/

[2] The source code can be accessed at https://github.com/reynoldsnlp/flair. A local instance can be easily deployed using `docker-compose`.

features unique to Persian, adding each feature to a list of grammatical constructions for all supported languages. Each grammatical feature is added to a menu that allows the user to rank and sort each text based on the listed features.

Adding back-end support and text processing for Persian required implementing a new NLP pipeline. Previously supported languages (English, German, Russian, and Arabic) rely on Stanford CoreNLP, which does not currently have models for the Persian language. We implemented Stanza (Qi et al., 2020) as a separate API[3], which includes a Persian model based on the Persian Universal Dependency Treebank (Seraji) model for Farsi tagging (Seraji, 2015; Seraji et al., 2016). This required designing a new text processing pipeline inside FLAIR to access the external Stanza API for grammatical parsing. Once the document is processed and tokenized, we are able to extract lemmas, part of speech, grammatical feature tags, and syntactic dependency relations. We then search the grammatical features of each word, attaching labels for identified grammatical constructions. For example, a token categorized as a verb then is searched for grammatical feature tags relating to form, tense, person, negation, and plurality. On the front end of the website, the user then chooses to prioritize documents containing auxiliary verbs and negated verbs, with a higher ranking weight on documents containing negated verbs. The program then sorts the documents and move those containing a higher concentration of the desired feature to the top of the list. In the example below, the query is ranked by texts containing auxiliary verbs and negated verbs. Each result represents a Persian webpage from a Bing search for the bolded term. Based on FLAIR's user-selected weighted rankings, this query ranks results on quantity of negated verbs. The selected article (the Persian Wikipedia page for the Quran) therefore has the highest quantity of negated verbs, despite other articles potentially containing more auxiliary verbs. (Fig. 1).

Figure 1. A FLAIR query for simple question words and phrases ('Who? What? Where?, bolded text at top of image).

The grammatical constructions included in the Persian module can be broken down into sentence types and parts of speech. Our evaluation of sentence types allows the user to sort by specific question words (who, what, where,

[3] Using https://github.com/lingmod-tue/stanza-api

when, why, how). We also use the presence of coordinating and subordinating conjunctions to label subordinate clauses and to label sentences as either compound or complex. The presence of relative pronouns is used to label relative clauses.

When sorting by parts of speech, the user is able to select verbs based on verb form (participle adjective, present participle, auxiliary, finite, infinitive, and negated verbs), person, (first, second, and third person, sorted by plural and singular verbs), and tense (simple past, past progressive, simple present, and simple future). Verbs are also sortable by mood (subjunctive and imperative). Quantifiers are sortable by the four most common words (any, some, none, many), as well as by all quantifier words. Adjectives are sortable by degree (positive, comparative, superlative). Adverbs can be categorized as either temporal, locational, negative, or other. Users can rank text by all pronouns, or specify demonstrative, indefinite, interrogative, negative, personal, reciprocal, reflexive, or relative pronouns. Numbers are identified and tagged as either cardinal or ordinal. Nouns are categorized as either plural or singular. Texts can also be ranked by total concentration of all prepositions.

4. Discussion and Conclusions

Persian's literary form has remained largely unchanged for hundreds of years (Jeremias, 2004). Most grammar books and available language learning resources focus on this formal, historical variety, often used in academic reports and traditional poetry. However, the Persian spoken colloquially by roughly 100 million individuals worldwide often bears little outside resemblance to the Persian of textbooks and pedagogy. Colloquial Persian, the language of everyday conversation, online discourse and political discussion, is constantly changing, adapting, and evolving. Mastery of this everyday Persian is absolutely critical for any student who wants to master the Persian language. Yet, when studying Persian, language learners frequently struggle to adapt to colloquial speech and text. The pronunciation and spelling of common verb endings, abbreviations of certain words, and contraction of grammatical markers unique to Persian all become problematic when language learners transition from literary Persian to colloquial texts.

FLAIR's unique text processing capabilities and search functions provide language learners with access to written Persian texts, both colloquial and formal, on any number of desired topics. FLAIR then presents each of these texts through a grammatical lens, showcasing the grammar of the Persian language in a more interactive and engaging way. Beginners can use Persian FLAIR to view simple grammatical constructions and features, first observed in the classroom, in texts that would have otherwise been too difficult for learners at their level. Learners at intermediate and advanced levels can seach vast repositories of online information and find level-appropriate texts, sorting and filtering the texts based on grammatical features they are working to master. FLAIR allows users to turn texts of any length, topic, intended audience, or level of formality into a personalized, fully customizable grammar textbook.

Over time, we hope to augment our work on Persian FLAIR to include more grammatical relations, using the dependency tags in the Seraji tree bank to allow the program to examine complicated syntactic relationships in depth. Users can see how each individual word in the sentence relates to the others, highlighting in-context usage of prepositions, grammatical markers, and possessive indicators unique to Persian and how they interact with other aspects of the language. In the future, we also hope to utilize FLAIR's innate localizability to provide native Persian speakers with access to FLAIR's functionality in the currently supported languages of English, German, Russian, and Arabic. Additionally, in future updates, we hope to implement a crowdsourced, machine learning based Persian readability assessment designed by Mohammadi & Khasteh (2020) to more accurately assess text difficulty and give users better rankings.

In conclusion, we have implemented a new Persian module in FLAIR, which allows teachers and learners to filter web search results for 56 grammatical constructions based on lexemes, parts of speech, and other morphosyntactic features. This module will be a useful addition to the Persian teacher's classroom tools, expanding access to this language and its rich history. Persian FLAIR will allow teachers to create more engaging and effective lessons by providing them with a way to target specific grammatical concepts. It will also

help learners to improve their understanding of Persian grammar by giving them access to a wide range of examples. We believe that this will be a valuable resource for both teachers and learners of Persian. We hope that it will help to promote the study of this beautiful and important language, expanding access to the language itself and to its rich history.

Acknowledgements

We would like to acknowledge funding from the Brigham Young University Department of Humanities, Office of Digital Humanities.

References

Chinkina, M., Kannan, M., & Meurers, D. (2016, August). Online information retrieval for language learning. In *Proceedings of ACL-2016 System Demonstrations*, 7–12.

Chinkina, M., & Meurers, D. (2016, June). Linguistically aware information retrieval: Providing input enrichment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, 188–198.

Dayani, M. (1990). A criteria for assessing the Persian texts' readability. *Journal of Social Science and Humanities*, 5(2): 35–48.

Gass, S., & Varonis, E. (1994). Input, interaction, and second language production. In *Studies in second language acquisition*, 16(03):283–302.

Hveem, J. (2019). RAFT: Readable Arabic Finding Tool. Unpublished masters project manuscript, Department of Instructional Psychology and Technology, Brigham Young University, Provo, Utah. Retrieved from https://scholarsarchive.byu.edu/ipt_projects/22

Jeremias, E. M. (2004). "Iran, iii. (f). New Persian". *Encyclopaedia of Islam*. Vol. 12 (New Edition, Supplement ed.). p. 432.

Krashen, S. (1977). Some issues relating to the monitor model. *On Tesol*, 77 (144–158).

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 55-60.

Mohammadi, H., & Khasteh, S. H. (2020, August). A machine learning approach to Persian text readability assessment using a crowdsourced dataset. In *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 1–7.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Reynolds, R. (2022). *FLAIR (Russian and Arabic)*. Office of Digital Humanities. https://odh.byu.edu/projects/flair-russian-and-arabic/

Schmidt, Richard W. (1990). The role of consciousness in second language learning. *Applied linguistics*, 11(2):129–158.

Seraji M. (2015). Morphosyntactic Corpora and Tools for Persian. Doctoral dissertation. *Studia Linguistica Upsaliensia*, 16.

Seraji M., Ginter, F., & Nivre, J. (2016). Universal Dependencies for Persian. In *Proceedings of the 10th International Conference on Language Resources and Evaluation* (LREC 2016), 2361–2365.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In *Input in second language acquisition*, 15:165–179.

Windfuhr, G. (2009) *The Iranian Languages*, Routledge 2009, p. 418.

EUROCALL 2023: CALL for all Languages
15–18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16945

Are digital learning paths sufficient for learning/practising English grammar? Views of secondary school pupils

Eirini Busack[a]
[a]English Department, Karlsruhe University of Education, eirinibusack@gmail.com

How to cite: Busack, Ei. (2023). Are digital learning paths sufficient for learning/practising English grammar? Views of secondary school pupils. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16945

Abstract

This study aims to investigate 80 grade 6–10 secondary school English as a Foreign Language (EFL) learners' opinions in Germany about the extent to which they consider Digital Learning Paths (DLPs) sufficient for acquiring or refreshing their English grammar knowledge. The learners were given one to two grammar-focused DLPs, depending on their grade and language level, to complete during their English classes. Upon completion, learners participated in an online survey. Research data were collected in October 2022 using an online survey. Interestingly, results from 80 learners showed that for half of the respondents, the DLPs were sufficient to learn or practise English grammar, but the second half of the respondents indicated that they would prefer DLPs to be integrated into grammar lessons at school. This is because these learners prefer to have additional explanations from the teacher. This case study shows that it can help promote DLPs as an alternative teaching method and encourage EFL teachers to deliver their English grammar lessons in a more engaging and sustainable format.

Keywords: *autonomous learning, English grammar, EFL, secondary school learners, digital learning paths.*

1. Introduction

This study aims to determine grade 6–10 secondary English learners' perceptions on the extent to which they consider DLPs sufficient for acquiring or refreshing their English grammar knowledge. An extensive online literature search found no research on the integration of DLP in the teaching of English grammar in secondary education. As part of the project *Sustainable Integration of Digital Teaching and Learning Concepts* (InDiKo, 2020-2023), the English Department at the Karlsruhe University of Education, Germany, developed the concept of teaching English grammar through DLPs. In particular, the English Department designed the seminar, *Media Didactic Competence: Learning Paths and Digital Storytelling for Grammar Teaching*, in which pre-service English teachers learn how to construct DLPs to teach grammar more interactively and effectively. These DLPs are then taken to secondary schools to be tested with English learners.

In many contexts, it has become common for learners to use tablets in class (Bjørgen et al., 2021) and do their homework online (Magalhães et al., 2020). Suddenly, however, the outbreak of the COVID-19 pandemic in early 2020 urged instructors, policymakers, and education institutions to acknowledge the importance of developing new concepts for using digital technology and virtual learning platforms to support distance learning. Many

education institutions worldwide have started implementing "Emergency Remote Education" (ERE) (Hodges et al., 2020) by using asynchronous learning and remote synchronous delivery, a combination of synchronous and asynchronous teaching formats.

The sudden need to adapt to online teaching and learning due to the COVID-19 outbreak challenged teachers' and learners' digital literacy, as the majority were unfamiliar with the concept and nature of distance learning (Crawford et al., 2020). Such a situation often results in learners being unable to engage with the asynchronous learning materials, which can lead to them being unable to follow the course (Bedenlier et al., 2021).

Given the fact that pupils reacted positively towards the DLP-based English grammar lessons, according to our previous studies (Busack, 2023), it was decided to look more specifically into the preferences of today's generation of "digital natives" (Prensky, 2001), i.e. to find out whether learners prefer to learn English grammar completely online or with live synchronous input and interaction with the teacher.

In particular, the focus is on DLPs that are built on the six levels of Bloom's Taxonomy, namely, recall previously learned information, understand new facts, apply knowledge to current situations, analyse ideas into simpler parts, find evidence to support generalisations or evaluate information to construct alternative solutions, and produce artifacts. Teachers who use Bloom's Taxonomy aim to help learners develop higher-order thinking skills by assisting them in practising their lower cognitive skills.

The innovative aspect of DLPs lies in their multimodality, i.e. their ability to cater for multiple learning styles and language levels simultaneously, and finally, their ability to provide ubiquitous learning in a safe, teacher-controlled learning space. As for the role of digital storytelling in DLPs, it is used to stimulate learners' curiosity and create a more entertaining and participatory language learning experience.

This study intends to fill in the aforementioned gap in the literature and to answer the following research question: are DLPs sufficient for secondary English learners to learn/practice grammar, or do learners prefer the DLP to be part of the grammar lessons, together with other materials from the teacher?

2. Method

2.1 Context and participants

The DLP-based grammar lessons were conducted at a secondary school in Baden-Württemberg, Germany. In terms of context, the school teachers chose one of the grammar-focused DLPs that our pre-service teachers had prepared and combined it with their own learning materials and activities (e.g. pre-DLP activity and post-DLP activity) to deliver the grammar lesson. The pre-DLP activity usually consisted of questions to help English learners familiarise themselves with the content of the DLP, while the post-DLP activity usually consisted of a task where pupils had to use the specific grammar phenomenon to produce an artifact or output, such as a letter, diary entry, etc. Each lesson lasted 60 minutes and took place in the school computer room. Each learner was provided with a computer and a headset. At the end of the lesson, learners were asked to complete an anonymous online survey to give their impressions.

The sample included 80 EFL learners with an age range of 11-15 years. These learners were studying in the grades 6, 7, 8, 9, and 10. There were 47 boys and 33 girls with different language proficiency levels and learning styles. They have been studying English at school between three and seven years.

In this particular school, English grammar is taught primarily by books and non-digital materials supplied by the teachers, resulting in limited ways to accommodate students' diverse learning preferences and abilities.

2.2 DLP material and procedure

A DLP is an internet-based learning path that provides a series of coordinated tasks through interactive materials that can help learners develop the habit of independent and self-directed learning while working towards a specific learning goal. In addition to visuals, it is also possible to use audio-visuals to support learning (Windler & Wolf, 2021).

According to Schmidt (2009), DLPs' ultimate goals are; 1) knowledge acquisition in terms of completing a transferability task and applying the newly gained knowledge in a related day-to-day exercise example, and 2) cognition in the sense of being able to reflect on one's own learning, i.e. on the connections between the previously acquired knowledge and ideas used in reasoning to reach the acquisition of the new piece of knowledge.

DLPs, as long as they are embedded in a pedagogically meaningful way, can be a supportive tool for learners both in the classroom and at home (Schmidt, 2009). They can include Word, Excel, and PowerPoint documents, pictures, web pages, links, memos, and activity items, such as polls, tests, quizzes, tasks, discussions, etc. (Busack, 2023). The DLPs of our pre-service teachers included interactive grammar exercises like true/false, fill-in-the-blanks, matching exercises, etc. Learners could access the grammar activities according to their level (basic, intermediate, advanced) based on their results from previous exercises, using an automated marking system, or they could freely choose the activities they wanted to do based on their perceived proficiency level. The formative assessment feature allowed learners to review their answers and receive feedback anytime. Furthermore, our DLPs presented grammar theory multimodally, such as images, videos, and audio files. An educational agent (non-moving avatar or character) was also present to make the DLP more attractive. The DLP topics included: conditional clauses 0, 1, 2, and 3; relative clauses; simple present and present progressive, will future and future continuous, simple past, present perfect, past perfect, and future perfect tenses; reported speech; and active and passive voice. The DLPs created by our pre-service English teachers are hosted at the University's learning management system *Innovation Space*. To see how our DLPs look in action, please scan the QR code in the appendix. These DLPs are structured upon Bloom's Taxonomy, a useful checklist to ensure that each level of grammar learning is assessed and appropriate assessment techniques are used in the classroom.

2.3 Data collection and analysis

As we wanted to collect responses from a large number of participants, we decided to use a quantitative research design, so the instrument used to collect the data was an anonymous online questionnaire given to each English learner at the end of the DLP-supported English grammar lesson. The anonymity was intended to help participants give honest feedback. The questionnaire consisted of 15 closed- and open-ended questions. The data were analysed using the statistical software Jamovi.

3. Results

Answers in the online survey surprisingly revealed that 40 (50%) English learners would prefer to learn or practise English grammar autonomously through DLPs, while the other 40 (50%) English learners stated that it would be better for them to integrate the DLPs into the English grammar lessons at the same time while benefiting from teacher's presence and their explanations (as demonstrated in Figure 1).

Further responses to the survey revealed that the pupils enjoyed the formative assessment function and the ability to check their answers instantly. They also found the learning agents to be a great idea as they were considered learning aids with the narrative background that the lesson follows, making the otherwise dull grammar lesson engaging and student-centered. Furthermore, pupils appreciated the interactive exercises and the DLPs that used videos and comics in their content. Several responses from the questionnaire are listed below:

- "It's very important for me that I get a direct explanation of what went wrong."
- "Getting positive comments for my correct answers motivates me."
- "It's very interactive."
- "It's motivating and makes learning not seem so lonely and boring."
- "I like DLPs because you can interact directly with exercises. "
- "I like that you get to do something yourself."

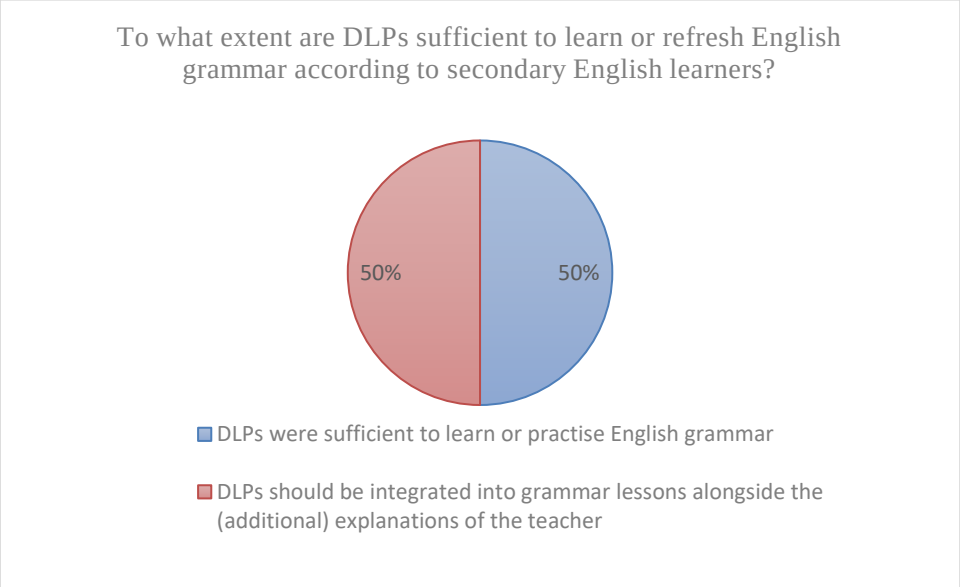


Figure 1. The extent to which DLPs are sufficient for the learning or refreshing of English grammar according to secondary English learners.

4. Discussion and conclusions

This study explored how far DLPs are considered sufficient for learning/refreshing English grammar knowledge by secondary school EFL learners. The study's findings imply that our sample is torn between having DLPs as the only way of learning/practising English grammar, and having DLPs as part of the English grammar lessons but combining them with learning materials and explanations provided by their English teacher.

The study moreover found that English learners generally appreciated the fact that using DLPs allowed them to check and correct their answers anytime, without waiting for the teacher to find time to correct them, and DLPs' scenario-based learning feature, which made them curious about the end of the story and willing to go through the exercises to find out more about the lesson.

In summary, these findings suggest that even though the current generation of learners are known as "digital natives" (Prensky, 2001), it cannot be taken for granted that they are all equally comfortable with autonomous digital learning. On the one hand, some learners may be more autonomous and want to take charge of their learning by exploring the given digital context. On the other hand, other learners would still need guidance in their learning and therefore seem not to welcome those parts of lessons that are being delivered asynchronously. Based on the finding that 50% of our sample decided that it would be better to combine DLPs and teacher's simultaneous explanations, this may be due to students' traditional mindset and difficulty in adapting to self-paced and digital learning. In this sense, the learners may think that having the teacher always present when trying to practise or learn a (new) concept makes them feel safe in their learning. Ultimately, based on the result of the first 50% of our sample, it can be assumed that pupils who have a positive attitude towards online learning will be more enthusiastic about using self-access and virtual learning materials than those who do not prefer online learning.

A further study is planned as an extension, teaching English grammar only via DLPs over a six-month period and seeing how grade 6-10 EFL learners' attitudes and academic success develop. Future studies could involve participants from other schools in different countries to contribute to the comparability and generalisability of the results. Moreover, conducting a study to combine surveys with structured interviews (mixed methods) to better

understand the relationships or inconsistencies between the qualitative and quantitative data in the surveys could help make the findings more generalisable.

Acknowledgments

This project is part of the 'Qualitätsoffensive Lehrerbildung', a joint initiative of the Federal Government and the Länder, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research in Germany. The authors are responsible for the content of this publication.

Appendix

Guest key: 3juqYON&z@QHkg;&97Is

References

Bedenlier, S., Wunder, I., Gläser-Zikuda, M., Kammerl, R., Kopp, B., Ziegler, A., & Händel, M. (2021). Generation invisible?. Higher Education Students' (Non)Use of Webcams in Synchronous Online Learning. *International Journal of Educational Research Open 2*(100068). https://www.sciencedirect.com/science/article/pii/S2666374021000388

Bjørgen, A. M., Fritze, Y., & Haugsbakk, G. (2021). Dealing with increased complexity. Teachers' reflections on the use of tablets in school. *Pedagogies: An International Journal*, 1–16. Retrieved from: https://www.tandfonline.com/doi/full/10.1080/1554480X.2021.1897010

Busack, Ei. (2023). THE POTENTIAL OF DIGITAL LEARNING PATHS FOR PROVIDING EQUITABLE CONDITIONS FOR LEARNING ENGLISH GRAMMAR. In E. A. Sheehan & M. Köhler (Eds.), *Proceedings of the 9th International Conference on Education* (pp. 83-94). TIIKM Publishing. https://tiikmpublishing.com/proceedings/index.php/icedu/article/view/1169

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P.A., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, *3*, 1–20. https://journals.sfu.ca/jalt/index.php/jalt/article/view/191

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*, *27*, 1–12. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Magalhães, P., Ferreira, D., Cunha, J., & Rosário, P. (2020). Online vs traditional homework: A systematic review on the benefits to students' performance. *Computers & Education*, *152*, 103869. https://www.sciencedirect.com/science/article/pii/S0360131520300695?via%3Dihub

Pädagogische Hochschule Karlsruhe. (n.d.). *Nachhaltige Integration von fachdidaktischen digitalen Lehr-Lern-Konzepten an der PH Karlsruhe (InDiKo, 1.05.2020 - 31.12.2023).* Retrieved July, 08, 2023 from https://www.ph-karlsruhe.de/projekte/indiko.

Prenksy, M. (2001). Digital natives, digital immigrants. *On the Horizon, 9*(5), 1-6. https://www.marcprensky.com/writing/Prensky%20%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf

Schmidt, R. (2009). *Selbstgesteuertes Lernen durch Lernpfade* [Paper Präsentation]. 100. MNU Kongress, Regensburg, Germany. https://www.digitale-lernpfade.de/theorie/Selbstgesteuertes%20Lernen%20durch%20Lernpfade.pdf

Windler, M., & Wolf, K. (2021). Entwicklung und Erprobung digitaler Lernpfade für den Mathematikunterricht in heterogenen Klassen. In K. Hein, C. Heil, S. Ruwisch & S. Prediger (Eds.), *Beiträge zum Mathematikunterricht 2021*. WTM Verlag. http://dx.doi.org/10.17877/DE290R-22338

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16918

Effective conversational practice in an Icelandic LMOOC

Kolbrún Friðriksdóttir[a]
[a]Faculty of Icelandic and Comparative Cultural Studies, University of Iceland, , kolbrunf@hi.is

How to cite: Friðriksdóttir, K. (2023). Effective conversational practice in an Icelandic LMOOC. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik.
https://doi.org/10.4995/EuroCALL2023.2023.16918

Abstract

This paper presents a course design that aims to effectively integrate conversational practice into an Icelandic Language Massive Open Online Course (LMOOC). Additionally, it introduces an ongoing survey study investigating the perceptions and engagement of learners (n = 60) in live oral communication sessions that are offered within the program's specific distance learning mode. The study aimed to: a) ascertain whether students are motivated to participate in the optional group sessions; b) identify the main factors that motivate them to participate in these sessions; c) determine whether the live sessions fulfill their expectations; and d) identify why some students choose not to attend the live sessions offered. Results indicate that most of the participants in the study took advantage of the opportunity and attended some or all of the conversation classes offered. The results also show that attendees were primarily motivated by the opportunity to practice communication and pronunciation in the target language. Moreover, the findings demonstrate that most of the participants in both courses highly valued the live sessions, finding them to either meet or surpass their expectations. Regarding non-attendees, data suggests that scheduling constraints and anxiety about speaking and meeting others posed challenges for some individuals.

Keywords: *conversational practice, LMOOCs, distance learning mode, human tutor support, Icelandic Online.*

1. Introduction

Because they are open to an unlimited number of participants, LMOOCs pose great challenges in technical, pedagogical, and practical terms when it comes to providing speaking opportunities for L2 learners. Drawing upon existing literature, several LMOOCs have experimented with diverse approaches to increase oral communication within these learning environments. Gimeno-Sanz (2017) reported on an LMOOC where students initiated spontaneous interactions and peer support naturally emerged. This led the course designers to organize both instructor-led sessions and learner-driven speaking practice sessions in Google Hangouts. Appel and Pujolà (2021) implemented an eTandem language learning approach, including tutors' support, to incorporate speaking interaction within LMOOCs. That design enables students to develop the necessary competences for engaging in online synchronous speaking interaction with native/near-native speakers of their

target languages. Bárkányi (2021) demonstrates the integration of speaking practice in LMOOCs through the utilization of voice recording tools, where students share links to their recordings in discussion forums and invite instructors and peers to provide feedback. A recent critical review of LMOOC design features (Chong et al., 2022) noted that one of the primary limitations of LMOOCs is the lack of opportunities they afford for instructor-learner and learner-learner interaction.

This paper introduces student perceptions of a course design aimed at effectively integrating conversational practice into the LMOOC program Icelandic Online (IOL) (https://icelandiconline.com) by incorporating a virtual social space for both instructor-learner and learner-learner interactions. The paper presents a survey study focused on learners' experiences and usage of the specific oral communication sessions that are now provided within the program. IOL was developed at the University of Iceland for second/foreign-language learners of Icelandic and comprises comprehensive and organized online language learning materials. The program provides seven free and open consecutive courses that are self-guided and interactive. IOL's learning material can be accessed through a variety of learning modes, all of which are based on self-directed asynchronous learning: an open non-tutorial mode, a distance learning mode, and a blended learning mode. The distance learning mode, which is the focus of this study, was developed specifically to allow particular target groups in IOL to receive individual guidance and support from a human tutor and is run independently of the basic open massive course (Friðriksdóttir 2018, 2021a, 2021b). The two distance learning courses, IOL 1 (beginner level), and IOL 2 (lower intermediate level), are run for eight weeks at a time and are offered five times a year. These courses provide students with comprehensive web-based learning materials. Additionally, students are given the option to complete 12 written assignments that are embedded in the learning content, which are then reviewed and critiqued by a dedicated tutor. The courses finish with an optional online final exam. While the current IOL system is limited by its inability to provide effective conversational practice, the course content was recently expanded to include optional weekly hour-long live sessions where students can meet with a native instructor and peers in a virtual classroom on Zoom during the evening hours, Icelandic local time. Prior to the live sessions, students receive level-specific teaching materials closely aligned with the web-based course content they are also working on. These materials prepare students to actively engage in the upcoming sessions.

The following research questions guided this ongoing survey study: a) Did students enrolled in two different courses, IOL 1 and IOL 2, demonstrate motivation to participate in optional weekly group sessions for oral interaction in the target language?; b) what were the primary factors that motivated students to take part in these live sessions?; and c) did the sessions meet the participants' expectations? Also, the study sought to investigate d) why certain students did not take advantage of the opportunity to participate in the offered live sessions.

2. Method

To comprehend learner engagement and perceptions in the specific live oral communication sessions in focus, an anonymous post-course online questionnaire in English was distributed. The survey was sent to registered students (n = 152) enrolled in two distance courses, IOL 1 and IOL 2, offered between January 2021 and April 2023. The questionnaire included ten questions (most of them multiple choice) using a three- to five-point Likert scale. The survey items were primarily developed based on informal feedback obtained from previous learners' comments and complaints regarding specific elements (or lack of) of the IOL web-based learning materials. This feedback was collected through emails, learners' written assignments, as well as final exams administered within the web-based course (Section 1).

The target population included both those who had attended 1-8 live sessions over this period of time and also those who did not attend any live session but had been active in other aspects of their courses. A total of 60 students responded to the survey in full or in part: 37 in the IOL 1 course (out of 102 who were invited) and 23 in the IOL 2 course (out of 50 invited). Some of the students in IOL 2 had previously completed the IOL 1 course and had also attended the live sessions offered in that course. The respondent sample is varied and includes participants from three different continents. The largest age group in IOL 1 was 35-44 years old and in IOL 2 it was 45-54. Of the study population, over 60% of the participants in both courses were female. The

questionnaire incorporated a mix of closed-ended and open-ended questions. However, due to space constraints in this article, only a subset of the collected data will be presented. The unshared data comprises qualitative insights gathered by prompting learners to describe the most valuable aspects of the course's live sessions and suggest enhancements. Nevertheless, this paper provides a glimpse into this qualitative data in the following section.

3. Results

3.1 Participation in live sessions

Table 1 data reveals that the majority of participants attended 7-8 live sessions, with 48.7% in IOL 1 and 78.3% in IOL 2. A smaller percentage attended 4-6 sessions (16.2% in IOL 1 and 13% in IOL 2), while 18.9% in IOL 1 engaged in 1-3 sessions. Additionally, 16.2% in IOL 1 and 8.7% in IOL 2 did not attend any live sessions.

Table 1. Participation in live sessions in IOL 1 and IOL 2.

Total live sessions attended	IOL 1 (n = 37)	IOL 2 (n = 23)
7-8 live sessions	48.7% (n = 18)	78.3% (n = 18)
4-6 live sessions	16.2% (n = 6)	13% (n = 3)
1-3 live sessions	18.9% (n = 7)	0%
No live sessions	16.2% (n = 6)	8.7% (n = 2)

3.2 Motivating factors for attendance

With regards to the key motivating factors that prompted students to attend the live sessions, a significant proportion of participants in both IOL 1 and IOL 2 identified the opportunity to practice Icelandic conversation as the primary driving force. In IOL 1, 84% and in IOL 2, 90% recognized this as the main reason for their active participation, as shown in Table 2. Similarly, 80% in IOL 1 and 70% in IOL 2 valued practicing Icelandic pronunciation. Seeking assistance from a human tutor motivated 68% in IOL 1 and 70% in IOL 2. Additionally, the opportunity to seek guidance from the tutor on various aspects of the language was considered important by 40% in IOL 1 and 50% in IOL 2. Conversely, the aspect of connecting with fellow students in the course was deemed less significant, with only 20% of respondents in IOL 1 and 15% in IOL 2 identifying it as an important factor for attendance.

Table 2. Main reasons for attending the live sessions in IOL 1 and IOL 2.

Motivating factors	IOL 1 (n = 25)	IOL 2 (n = 19)
To practice Icelandic conversation	84% (n = 21)	90% (n = 18)
To practice Icelandic pronunciation	80% (n = 20)	70% (n = 14)
To get assistance from a human tutor	68% (n = 17)	70% (n = 14)
To ask the tutor about different aspects of the language	40% (n = 10)	50% (n = 10)
To meet other students in the course	20% (n = 5)	15% (n = 3)

3.3 Participants' expectations

As for the question of whether the sessions met participants' expectations, Table 3 shows that 72% of the learners in IOL 1 and 90% in IOL 2 reported that the live sessions either met or surpassed their expectations. In the IOL 1 course, 24% indicated that the sessions fell below their expectations.

Table 3. Categories of participants' expectations in IOL 1 and IOL 2.

Participants' expectations	IOL 1 (n = 25[1])	IOL 2 (n = 20)
Exceeded expectations	24% (n = 6)	25% (n = 5)
Met expectations	48% (n = 12)	65% (n = 13)
Below expectations	24% (n = 6)	0%

3.4 Demotivating factors for attendance

Table 4 presents insights based on data from the eight survey respondents who indicated that they did not attend any of the live sessions. The primary reason cited by both groups for not attending the live sessions was timing. This reason was expressed by both of the students in IOL 2 and by 66.7% of the participants in IOL 1. Furthermore, three of the IOL 1 learners expressed that anxiety related to speaking Icelandic and socializing with fellow students posed challenges. Additionally, one IOL 1 learner mentioned a preference for independent study.

Table 4. Main reason for not attending live sessions in IOL 1 and IOL 2.

Demotivating factors	IOL 1 (n = 6)	IOL 2 (n = 2)
The scheduled time of the live sessions was unfortunate for me	66.7% (n = 4)	100% (n = 2)
I was anxious about speaking Icelandic	33.3% (n = 2)	0%
I was anxious about meeting others	16.7% (n = 1)	0%
I prefer to study on my own	16.7% (n = 1)	0%

Finally, due to space limitations, only two quotes from learners are presented here to address the data obtained from the survey's open-ended questions asking about personal views towards the live sessions. On the one hand, a respondent explained the key takeaway from the course's sessions: "I learned how to communicate person to person and how to speak." On the other, another student suggested an improvement for the sessions: "It would be great to have an additional native speaker for small conversations."

[1] Note: One respondent in IOL 1 and two in IOL 2 chose the 'other (specify)' option and added comments.

Discussion

Based on the data presented, the majority of participants in the study took advantage of the opportunity and attended either all or some of the live sessions offered in the two LMOOC courses under investigation. The results reveal that the primary motivating factors for attendees were the chance to practice communication in Icelandic and improve their pronunciation. Furthermore, there is evidence suggesting that the availability of a human tutor who can answer questions and offer guidance regarding the target language also played a role in attracting learners to the live sessions. Additionally, the findings indicate that a significant number of participants in both courses held the live sessions in high regard, stating that they either met or exceeded their expectations. Regarding the non-attendees in the study, the data suggests that some individuals encountered challenges attending due to the timing of the live sessions. However, resolving this matter is not straightforward as the program consistently has students in multiple time zones around the world. In addition, a few non-attendees in the beginner course expressed social apprehension and anxiety about speaking the target language, widely recognized barriers that prevent numerous learners from fully engaging in speaking activities within language learning environments (Bárkányi, 2021).

4. Conclusions

In conclusion, this study contributes significant data on the design of an LMOOC that successfully integrates communication practice within the learning environment. It specifically highlights a tutored distance mode of delivery as a means of achieving this integration. Specifically, the study investigates how learners perceive and engage with live oral communication sessions provided in such a learning mode. This study's findings have practical implications for LMOOC designers and instructors, and offer ideas of effective integration of oral communication practice within such learning environments.

To overcome the limitations of the small sample size in the present study, future research should focus on conducting follow-up studies with larger sample sizes to gather data from a more extensive range of live sessions.

References

Appel, C. & Pujolà, J-T. (2021). Designing speaking interaction in LMOOCs: An eTandem approach, *ReCALL*, *33*(2), 161-176. https://doi.org/10.1017/S0958344021000045

Bárkányi, Z. (2021). Motivation, self-efficacy beliefs, and speaking anxiety in language MOOCs, *ReCALL*, *33*(2), 143-160. https://doi.org/10.1017/S0958344021000033

Chong, S. W., Khan, M. A., & Reinder, H. (2022). A critical review of design features of LMOOCs, *Computer Assisted Language Learning*, 1-21. https://doi.org/10.1080/09588221.2022.2038632

Friðriksdóttir, K. (2018). The impact of different modalities on student retention and overall engagement patterns in open online courses. *Computer Assisted Language Learning*, *31*(1-2), 53-71. https://doi.org/10.1080/09588221.2017.1381129

Friðriksdóttir, K. (2021a). The effect of tutor-specific and other motivational factors on student retention on Icelandic Online. *Computer Assisted Language Learning*, *34*(5-6), 663-684. https://doi.org/10.1080/09588221.2019.1633357

Friðriksdóttir, K. (2021b). The effect of content-related and external factors on retention in a LMOOC. *ReCALL*, *33*(2), 128-142. Special issue: Researching Massive Open Online Courses for language teaching and learning. https://doi.org/10.1017/S0958344021000069

Gimeno-Sanz, A. (2017). Designing a MOOC for learners of Spanish: Exploring learner usage and satisfaction. In Borthwick, K., Bradley, L., & Thouësny, S. (eds.), *CALL in a climate of change: Adapting to turbulent global conditions – short paters from EUROCALL 2017*. Dublin: Research-publishing.net, 122-127. https://research-publishing.net/manuscript?10.14705/rpnet.2017.eurocall2017.700.

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16967

Language teacher training in CALL: fostering engagement in an online learning environment

Giovanna Carloni[a]

[a]Department of Communication Sciences, Humanities and International Studies, University of Urbino, giovanna.carloni@uniurb.it

How to cite: Carloni, G. (2023). Language teacher training in CALL: fostering engagement in an online learning environment. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16967

Abstract

The present study aims to investigate how effective pre-service language teachers perceived the engagement fostered in an online language pedagogy course. The course was targeted at developing pre-service language teachers' content-specific knowledge and digital skills through CALL within a language teacher training program implemented at an Italian university. The instructor used various learning environments and teaching methodologies to foster engagement. The data for the study were collected through an online semi-structured questionnaire administered to 33 pre-service language teachers at the end of the course. Findings show that pre-service language teachers deeply valued the high degree of engagement fostered in the course, although some challenges emerged at the beginning.

Keywords: *teacher training, CALL, engagement, online learning.*

1. Introduction

University courses have moved online extensively in the last few years. However, engagement is still a challenge in online learning environments (Veletsianos, 2020; Bergdahl, 2022). The present study aims to analyze how effective pre-service language teachers (in the remainder of the paper, I will refer to the pre-service teachers as 'students' as they took part in the course in this capacity) perceived the engagement fostered in an online language pedagogy course targeted at developing students' content-specific knowledge and digital skills through CALL. The course was implemented within an online language teacher training program offered at an Italian university.

Since engagement with content, peers, and instructors is pivotal in online learning environments (Garrison et al., 2001; Garrison, 2017; Darby & Lang, 2019), the instructor designed a course architecture targeted at promoting effective engagement in the online language pedagogy course investigated. In particular, engagement was enhanced through social collaboration platforms, Flipped Learning (Brinks Lockwood, 2018; Kotska & Marshall, 2017), and Peer Instruction (Mazur, 1997; Dancy et al., 2016).

2. Method

2.1. Online language pedagogy course architecture

To make students feel safe and respected in the online course investigated, the instructor fostered students' social presence, which is "the ability of participants to identify with the community (e.g. course of study), communicate purposefully in a trusting environment, and develop inter-personal relationships by way of projecting their

individual personalities" (Garrison et al., 2001). To foster social presence, a paradigm of the Community of Inquiry model (Garrison, 2017), the instructor designed icebreakers targeted at helping learners connect with their peers on a personal level. Furthermore, the instructor provided students with consistent formative feedback aimed at making them feel valued as individuals since "[s]ocial presence […] arises when you and your learners experience an atmosphere of safe and open enquiry and mutual support in your class" (University of Waterloo et al., n.d.).

In the language pedagogy course investigated, social collaboration was instrumental in fostering highly engaging knowledge co-construction from a socio-constructivist perspective (Lantolf et al., 2015; Hampel, 2020). Flipped Learning and Peer Instruction contributed to the development of engagement (with content, peers, and instructor) and played a pivotal role in teaching presence, which is "the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes" (Anderson et al., 2001).

In keeping with Flipped Learning, students engaged with learning materials before class; students then delved into content-specific constructs critically during live classes. In particular, before live classes, students engaged in social annotation on the Perusall platform: "social annotation […] [is] a type of learning technology that enables the addition of notes to digital and multimodal texts for the purposes of information sharing, peer interaction, knowledge construction, and collaborative meaning-making" (Hodgson, Kalir, & Andrews, 2023). On Perusall, students read the assigned scientific articles, inserted their comments, commented on their peers' comments, and answered their peers' questions. During live classes, students co-constructed their knowledge through Peer Instruction. In particular, students first answered individually a multiple choice question focusing on the content studied on Perusall before class; then, in small groups, learners discussed the answers provided; afterwards, students answered individually the same multiple choice question they had answered previously; and finally, the instructor showed the results of the multiple choice question (the one answered after group discussion) and commented on them (Mazur, 1997). Besides promoting active knowledge building, Peer Instruction enabled the instructor to provide students with consistent formative feedback, which is pivotal in online learning (Garrison et al., 2001; Garrison, 2017; Darby & Lang, 2019).

After live classes, self-selected small groups worked online to create sections of a digitally-enhanced lesson unit collaboratively. Before live classes, the instructor provided each group with video feedback on the teaching materials developed. The instructor thereby provided students with formative feedback consistently in keeping with online pedagogy (Garrison et al., 2001; Garrison, 2017; Darby & Lang, 2019).

2.2. Participants and context

The study was conducted in an online language pedagogy course targeted at developing pre-service language teachers' language pedagogy and digital skills through CALL. The course was implemented within an online pre-service language teacher training program at an Italian university. The cohort consisted of 33 students.

Students filled in an online self-evaluation semi-structured questionnaire after each synchronous class. The questionnaire was course-tailored. Soon after each live class, students completed the questionnaire where they could also ask questions and express their needs to the instructor.

2.3. Research question

The present study aims to answer the following research question: how effective did students perceive the engagement fostered in the online language pedagogy course?

The self-evaluation questionnaire was targeted at monitoring students' learning process, needs, and wellbeing. The data collected after the first live class revealed, for example, that students found the amount of digitally enhanced activities assigned overwhelming. The instructor thus modified the course structure accordingly in keeping with a pedagogy of care, which values students' wellbeing (Gleason & Mehta, 2022; Quinn et al., 2022).

2.4. Instruments and data collection

A mixed-method approach was used to carry out descriptive research. The data for the study were collected through an online semi-structured questionnaire that students filled in before the final exam (see Endnote). The questionnaire included two sections: (a) the validated Community of Inquiry questionnaire, which features a five-point Likert scale (where 1 stands for strongly disagree, 2 for disagree, 3 for undecided, 4 for agree, and 5 for strongly agree) (Caskurlu, 2018; Stenbom, 2018); and (b) course-tailored closed and open-ended questions, devised to identify students' perceptions on activity types and degree of engagement. The Community of Inquiry questionnaire is designed to identify students' perceptions on: teaching presence (design and organization, facilitation, and direct instruction); social presence (affective expression, open communication, and group cohesion); and cognitive presence (triggering event, exploration, integration, and resolution) (Caskurlu, 2018; Stenbom, 2018). In the present study, the Community of Inquiry questionnaire was used to investigate social presence and teaching presence.

3. Results and discussion

The data collected through the final questionnaire show that, in regard to pre-class activities, the majority of students (62.1%) found social annotation on Perusall especially suitable for fostering engagement. In particular, most students (75%) preferred reading their peers' comments; 42.9% of students preferred giving feedback to their peers, while 32.1% of students preferred writing comments. In terms of engagement during live classes, 99% of students found Peer Instruction very effective. Interestingly, as part of Peer Instruction, the majority of students (74.1%) highly valued answering multiple choice questions individually before and after group discussions.

Most students claimed that the instructor scaffolded their discussions effectively, making them focus on relevant issues (66.7% strongly agreed and 21.1% agreed), while 12.1% were undecided (Table 1). Likewise, the majority of students held that the instructor managed to make them spot controversial content-specific issues successfully (63.6% strongly agreed and 27.3% agreed), while 9.1% were undecided.

Table 1. Teaching presence.

Teaching presence	1 Strongly disagree	2 Disagree	3 Undecided	4 Agree	5 Strongly agree
The instructor helped to focus discussion on relevant issues in a way that helped me to learn			12.1%	21.1%	66.7%
The instructor was helpful in identifying areas of agreement and disagreement on course topics in a way that helped me to learn			9.1%	27.3%	63.6%
The instructor was helpful in guiding the class towards understanding course topics in a way that helped me clarify my thinking			9.1%	30.3%	60.6%
The instructor helped to keep course participants engaged and participating in productive dialogue			3%	24.2%	72.7%
The instructor helped keep the course participants on task in a way that helped me to learn			6.1%	27.3%	66.7%
The instructor provided feedback that helped me understand my strengths and weaknesses relative to the course's goals and objectives			9.1%	21.2%	69.7%
The instructor provided feedback in a timely fashion			6.1%	15.2%	78.8%

Students claimed that the instructor scaffolded their critical thinking in relation to course content successfully (60.6% strongly agreed and 30.3% agreed), while 9.1% were undecided. Likewise, most students claimed that the

instructor managed to engage them in effective online discussions (72.7% strongly agreed and 24.2% agreed), while 3% were undecided. The effectiveness of the scaffolding provided was confirmed further when most students claimed that the instructor enabled them to focus on the activities provided successfully (66.7% strongly agreed and 27.3% agreed), while 6.1% were undecided. Overall, findings suggest that the pedagogical architecture devised to promote engagement through a combination of Flipped Learning and Peer Instruction worked effectively; students perceived the added value of engagement as a key paradigm of course design. Although students often mentioned that it was the first time they engaged actively online, they got used to the learning practice rather easily; students expressed their appreciation for active learning in the post-class self-evaluation questionnaires.

Most students found that the instructor feedback enabled them to identify facilitators and challenges in relation to the course aims effectively (69.7% strongly agreed and 21.2% agreed), while 9.1% were undecided. In addition, most students held that they received feedback promptly (78.8% strongly agreed and 15.2% agreed), which is instrumental in fostering social presence and effective engagement in online learning environments; 6.1% were undecided. In particular, all students except one claimed that the weekly video feedback was effective (53.8% strongly agreed, 42.3% agreed, and 3.9% were undecided).

The instructor promoted engagement also through a specific out-of-class collaborative activity; noticeably, every week small groups worked together to devise a section of a teaching unit. Interestingly, the data collected through the course-tailored close-ended questions featured in the final questionnaire show that 90.9% of students claimed creating a teaching unit collaboratively was the activity that helped them to develop their content-specific knowledge and digital skills the most. Furthermore, the majority of students (70.4%) held that co-constructing the teaching unit increased their motivation, which shows how deeply engagement and motivation are intertwined.

As previously mentioned, social presence is instrumental in fostering student engagement. Findings show that most students claimed that the activities in which they engaged, such as ice breakers, helped them feel like in-group members (72.7% strongly agreed and 21.2% agreed), while 6.1% were undecided (Table 2).

Table 2. Social presence (see Endnote).

Social presence	1 Strongly disagree	2 Disagree	3 Undecided	4 Agree	5 Strongly agree
The instructor actions reinforced the development of a sense of community among course participants			6.1%	21.2%	72.7%
Getting to know other course participants gave me a sense of belonging in the course		12.1%		39.4%	48.5%
I was able to form distinct impressions of some course participants		4%	3%	60.6%	27.3%
I felt comfortable conversing through the online medium			24.3%	45.5%	30.3%
I felt that my point of view was acknowledged by other course participants			21.2%	57.6%	21.2%
I felt comfortable disagreeing with other course participants while still maintaining a sense of trust		6.1%	30.3%	45.5%	18.2%
Online discussions helped me to develop a sense of collaboration			12.1%	42.4%	45.5%

Likewise, the majority of students held that getting to know their peers helped them feel as part of a cohesive group (48.5% strongly agreed and 39.4% agreed), while 12.1% did not agree. Almost to the same degree, most students claimed that they got to know their peers rather well (27.3% strongly agreed and 60.6% agreed), while 3% of the students were undecided and 4% disagreed. Most students held that they were at ease while interacting online (30.3% strongly agreed and 45.5% agreed), while 24.3% were undecided. Exactly to the same extent, the majority

of students felt at ease while interacting with their peers and discussing various topics online. These findings suggest that most students felt safe and visible as individuals in online interactions; however, a few students still faced some challenges in online discussions. Most students claimed that their peers valued and respected their opinions (21.2% strongly agreed and 57.6% agreed), while 21.2% were undecided. Most students held that they managed to keep a trustful relationship with their peers while voicing different opinions on course topics (18.2% strongly agreed and 45.5% agreed), while 30.3% were undecided and 6.1% disagreed. These findings suggest that having one's opinions acknowledged and disagreeing online were still an issue for some students; thus, to improve online interaction in the next iteration of the course, students will be taught explicitly how to acknowledge their peers' opinions and how to disagree while maintaining a trustful relationship. In general, however, students felt mostly comfortable to a very high degree in online interactions, which is a positive result. Finally, it is noteworthy that the majority of students held that they developed collaborative practices through online dialogical interactions (45.5% strongly agreed and 42.4% agreed), which is pivotal for engagement to occur in online learning environments effectively; 12.1% were undecided.

Findings show that students appreciated extensively the high degree of engagement fostered in the course. The pedagogical added value of engagement in terms of knowledge and skill development, which surfaced in the analysis extensively. Findings also suggest that the changes the instructor made to some activity structures, to foster students' wellbeing, was successful; in this respect, students' feedback highlighted how grateful they were for the way the instructor listened to their needs and acted accordingly.

A limitation of the study is that the sample size (=33) is not sufficient to generalize the findings of the study but the findings may be useful to other instructors planning similar courses.

4. Conclusions

The design of the online language pedagogy course proved successful in terms of engagement. Students perceived the added value of engagement as instrumental in promoting knowledge co-construction, skills development, and social presence. Furthermore, the student-centered approach implemented in the course scaffolded active learning effectively leading students to feel increasingly motivated. Finally, it is noteworthy that the adoption of a pedagogy of care contributed to students' wellbeing significantly.

In the future, a longitudinal study of students' perceptions of the effectiveness of the engagement fostered in various iterations of the online language pedagogy course will be carried out.

References

Anderson, T., Rourke, L., Garrison, D. R., Archer, W. (2001). Assessing teaching presence in a computer conference environment. *Journal of Asynchronous Learning Networks*, *5*(2), 1-17.

Bergdahl, N. (2022). Engagement and disengagement in online learning. *Computers & Education*, 188, 104561. https://doi.org/10.1016/j.compedu.2022.104561

Brinks Lockwood, R. (2018). *Flipping the classroom: what every ESL teacher needs to know*. Ann Arbor: University of Michigan Press.

Caskurlu, S. (2018). Confirming the subdimensions of teaching, social, and cognitive presences: A construct validity study. *Internet and Higher Education*, *39*, 1-12.

Community of Inquiry questionnaire, https://coi.athabascau.ca/coi-model/coi-survey

Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: the case of Peer Instruction. *Physical Review Physics Education Research*, 12, 010110. https://doi.org/10.1103/PhysRevPhysEducRes.12.010110

Darby, F., & Lang, J. M. (2019). *Small teaching online: applying learning science in online classes*. San Francisco: Jossey-Bass.

Garrison, R. (2017). *E-learning in the 21st century: a community of inquiry framework for research and practice*. New York: Routledge.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7-23.

Gleason, B., & Mehta, R. (2022). Editorial. A pedagogy of care: Critical humanizing approaches to teaching and learning with technology. *Italian Journal of Educational Technology*, *30*(1), 4-17. https://doi.org/10.17471/2499-4324/1278

Hampel, R. (2020). *Disruptive technologies and the language classroom. A complex systems theory approach*. Cham, Switzerland: Palgrave MacMillan.

Hodgson, J., Kalir, J., & Andrews, C.D. (2023). Social annotation: promising technologies and practices in writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt & A. Shibani (Eds.), *Digital writing technologies in higher education* (pp. 141-155). Cham: Springer. https://doi.org/10.1007/978-3-031-36033-6_9

Kotska, I., & Marshall, H. W. (2017). Flipped learning in TESOL: past, present and future. In J. Perren, K. Kelch, J. Byun, S. Cervantes & S. Safavi (Eds.), *Applications of CALL theory in ESL and EFL environments* (pp. 223-243). Hershey, PA: IGI Global.

Lantolf, J., Thorne, S. L., & Poehner, M. (2015). Sociocultural theory and second language development. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 207-226). New York: Routledge.

Mazur, E. (1997). *Peer instruction: a user's manual*. Upper Saddle River, NJ: Prentice Hall.

Perusall, https://www.perusall.com

Quinn, J., Burtis, M., & Jhangiani, S. (2022). *Designing for care*. Hybrid Pedagogy Inc. https://pressbooks.pub/designingforcare

Stenbom, S. (2018). A systematic review of the Community of Inquiry survey. *Internet and Higher Education, 39*, 22-32.

University of Waterloo, Queen's University, University of Toronto, & Conestoga College (n.d.). *High quality online courses. How to improve course design & delivery for your post-secondary learners*. Pressbooks. https://ecampusontario.pressbooks.pub/hqoc

Veletsianos, G. (2020). *Learning online: the student experience*. Baltimore, MD: Johns Hopkins University Press.

Endnote

Community of Inquiry questionnaire, https://coi.athabascau.ca/coi-model/coi-survey

EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16969

Teacher training perspectives for virtual exchange: Initial actions from the E-LIVE European project

Ciara R. Wigham[a] and Kristi Jauregi-Ondarra[b]

[a]Activité, Connaissance, Transmission, éducation, Université Clermont Auvergne, , ciara.wigham@uca.fr and [b]Utrecht University, , k.jauregi@uu.nl

How to cite: Wigham, C.R.; Jauregi-Ondarra, K. (2023). Teacher training perspectives for virtual exchange: Initial actions from the E-LIVE European Project. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16969

Abstract

The E-LIVE project (Engaging Languages in Intercultural Virtual Exchange) is a three-year project funded by Erasmus+ Cooperation partnerships in school education (KA220-SCH). In this short paper, we report on two of the project's initial teacher training actions. Firstly, a training needs analysis conducted via questionnaire and focus groups with 39 in-service L2 teachers from associate partner schools (nine primary and 30 secondary). We report on felt needs and anticipated training needs and three key project actions that the needs analysis informed: teacher voices workshops, expert webinars, and virtual exchange coaching. Secondly, we describe a virtual exchange conducted between trainee teachers enrolled in French, Dutch, and Colombian teacher training institutions. The pedagogical design of the virtual exchange is detailed and two of the final task productions showcased. Post-virtual exchange questionnaire and focus group data allow us to report on the trainee teachers' perceptions of the virtual exchange activities. The paper concludes with a discussion of how to ensure the success of future virtual exchange initiatives based on the analysis of the questionnaire and focus group data.

Keywords: *virtual exchange, school education, teacher training, European projects, needs analysis.*

1. Introduction

Virtual Exchange (VE) is a linguacultural learning-teaching practice that fosters meaningful exchanges between learners in geographically distant locations and from diverse linguacultural backgrounds through internet-based tools and innovative online pedagogies (O'Dowd, 2018). Telecollaboration or virtual exchange projects have gained significant prominence in the field of education since the onset of the 21st century, as demonstrated by the European projects summarized in Appendix A. Notably, a significant portion of these initiatives has concerned university-level education. E-LIVE (Engaging Languages in Intercultural Virtual Exchange) is a three-year project (2021-2024) funded by the Erasmus+ Cooperation partnerships in school education (KA220-SCH) involving a team of specialists in TELL and 32 associate partner schools and teacher training institutions (E-LIVE, 2022). In line with the Digital Education Action Plan (DG EAC, 2021), the key project objectives are to:

- enrich language teacher training programmes in higher education to develop digital literacy skills and organisational, pedagogical, and intercultural competences of trainee-teachers by mainstreaming VE within teacher-training programmes; and

- make L2 teaching programmes at primary and secondary schools more meaningful by offering learners opportunities to engage in VEs.

This paper reports on two of the project's initial teacher training actions: a training needs analysis conducted with in-service L2 teachers from the associate partner schools and a VE conducted between French, Dutch, and Colombian teacher training institutions.

2. Teacher training needs

At the project outset in the Summer of 2022, a questionnaire and a series of focus groups were conducted to better understand the teaching beliefs, experiences, conditions, and training needs (related to VE pedagogy, digital literacy skills, and intercultural mediation) of a heterogeneous group of 39 language teachers from five European countries collaborating with the E-LIVE project as associate partner schools, and who would partake in VEs in Summer-Autumn 2023. Nine primary-level educators responded to the questionnaire and 30 teachers at secondary-level. Twenty participants went on to partake in focus groups.

The full needs analysis is detailed in Clavel-Arroitia et al. (2023). Due to space limitations, we summarise emerging themes.

The needs analysis underlined current teaching challenges including integrating multilingual pupils in the L2 classroom and teaching for different competence levels within the same class. These challenges, or felt needs, were irrespective of teaching level, differences in student population, or teaching experience. On a five-point scale, ranging from 1 'not at all challenging' to 5 'extremely challenging', 50% of all participants ranked 'coping with differences in student population and enhancing inclusion' at 4 or above, and 14/30 secondary school teachers and 5/9 primary teachers ranked 'enhancing meaningful communication processes' at 4 and above (see Appendix B). In comparison, 33% of secondary school teachers ranked 'addressing cultural issues, developing intercultural competences and European citizenship' as very challenging (7/30) or extremely challenging (3/30), and 33% of primary teachers as very challenging (3/9).

Regarding anticipated training needs, primary educators strongly expressed needs related to the technical aspects of VE both in terms of choosing digital tools and technical support during exchanges. Secondary teachers expressed insufficient knowledge concerning the pedagogical integration of digital tools in language education as the barrier to integrating VE in their programmes. Three sub-themes appeared: the logistics of setting up VE, task design, and technical aspects. Regarding preferred training formats, a focus on learning from other colleagues experienced in VE emerged with suggestions for novices to 'job-shadow' more experienced colleagues during VEs and peer discussions focusing on good practice. Indeed, several teachers described a need for a "shared space" in which to meet and exchange with other practitioners, potentially more experienced in VE:

> *What works is the teaching community – we can have others contribute from time to time but in our day-to-day professional lives it's our colleagues who influence us […] asking questions to another more experienced teacher: Do you think this could work? What do you think if I did this? How many sessions ahead should I do this? etc.* (Participant3, Focus groupFR1)

The outcomes of the needs analysis were taken into account to organise online 'teacher voices workshops' (E-LIVE, 2022b), in which experienced VE practitioners showcased different VE school projects. These workshops focused on motivation and engagement in VE, new technologies for VE include Frame VR, and addressing diversity and differentiation in VE. The needs analysis also informed the choice of themes covered during expert webinars held in Autumn 2022- Spring 2023, namely differentiation and lingua franca exchanges. For VEs that are set up between associate partner schools and in which the needs analysis participants will participate, E-LIVE has also adopted a coaching approach (cf. Jauregi & Melchor-Couto, 2017). Coaches work alongside the teachers to prepare, plan, carry out, and evaluate exchanges in their language courses in a collaborative manner. A questionnaire in Summer 2023 will evaluate the effectiveness of this approach.

3. Teacher training virtual exchange project

An initial VE teacher-training project brought together trainee teachers (n=48) of foreign languages (English, Dutch, French, German, Spanish) from Utrecht University (NL), Université Clermont Auvergne (FR), and Pontificia Universidad Javeriana (CO). At Utrecht University, the virtual exchange was integrated into the Master's course, 'Language Curriculum'. At Université Clermont Auvergne, an 18-hour course entitled 'Training for pedagogical innovation - telecollaboration' was created within the didactics module of a Master's in English language teaching. Participants from Pontificia Universidad Javeriana were graduates within the Teacher education programme. The VE ran over a period of six weeks in Autumn 2022. The overall aim was to introduce VE and VE task design. After theoretical introductions in each institution in which cohorts interacted with training materials produced in the E-LIVE project relating to digitalisation and language education, engaging pedagogies for VE and task design, and intercultural communication (Jauregi-Ondarra & Canto, 2023), the VE was organised around three tasks (see Figure 1).

The exchange was organised around three tasks that utilised the affordances of both synchronous and asynchronous tools (Flip [a video discussion and sharing app: https://info.flip.com/], Moodle, BigBlueButton videoconferencing, Genial.ly [app to design interactive content: https://genial.ly/], and Padlet [interactive multimodal board: https://padlet.com]) to introduce virtual exchange and virtual exchange task design to the trainee teachers.

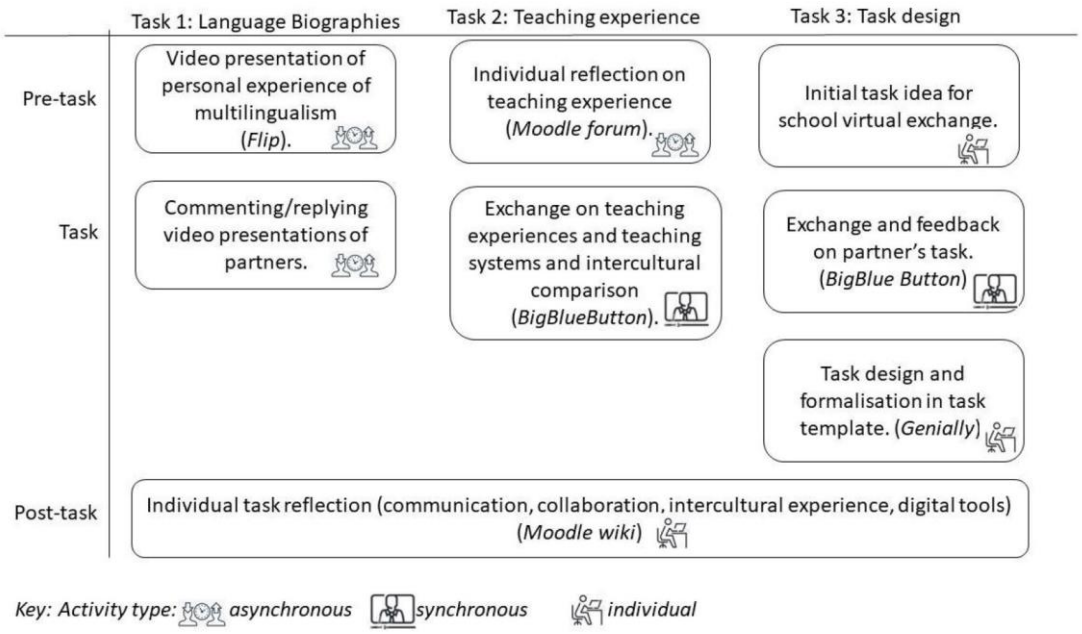


Figure 1. Pedagogical design of VE.

The final task (task 3) involved the trainees co-designing virtual exchange tasks that could be used for a class in their current teaching placement context (secondary schools). Trainees were asked to consider the learning objectives, environments and tools to be used, interaction patterns, and learner organisation as well as the expected production or outcome. These tasks were formalised using the E-LIVE task template in Genially and are currently being published within the project taskbank as open-access resources for other teachers interested in VE.

An example task produced, for a bilingual teletandem model of VE (Brammerts, 2002), was for learners to initially present one of their favourite cities then to research their VE partner's favourite city as regards accommodation, restaurant, and activity possibilities before exchanging these ideas with a partner and receiving feedback on which options they should choose before producing a written travel plan for which the L1 peer would provide linguistic feedback (see Figure 2).

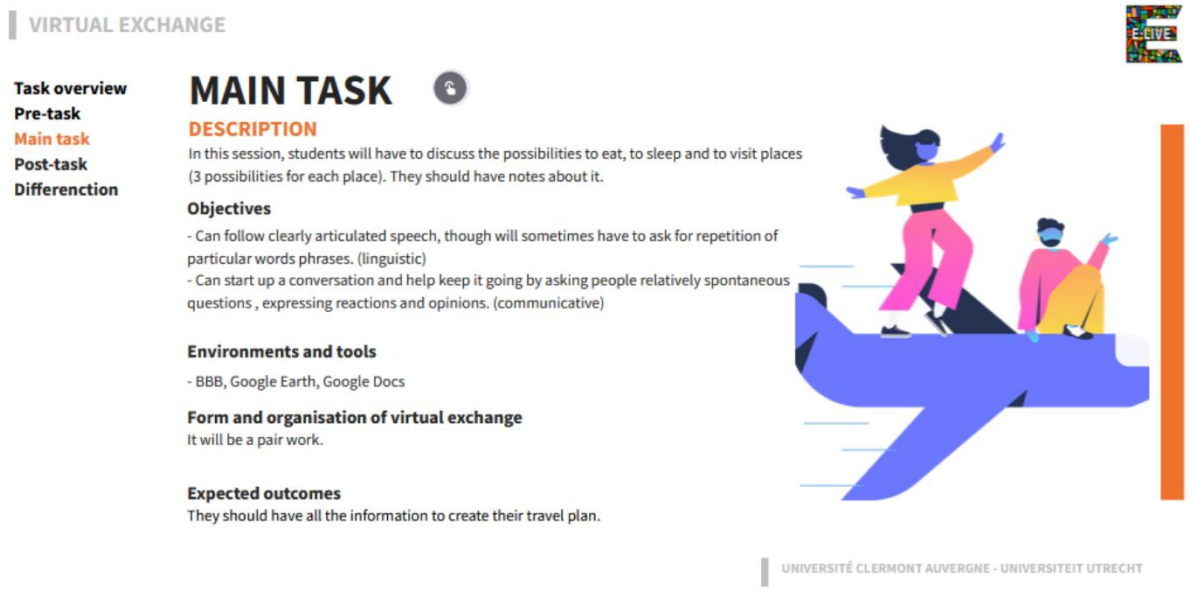


Figure 2. Sample task production 1.

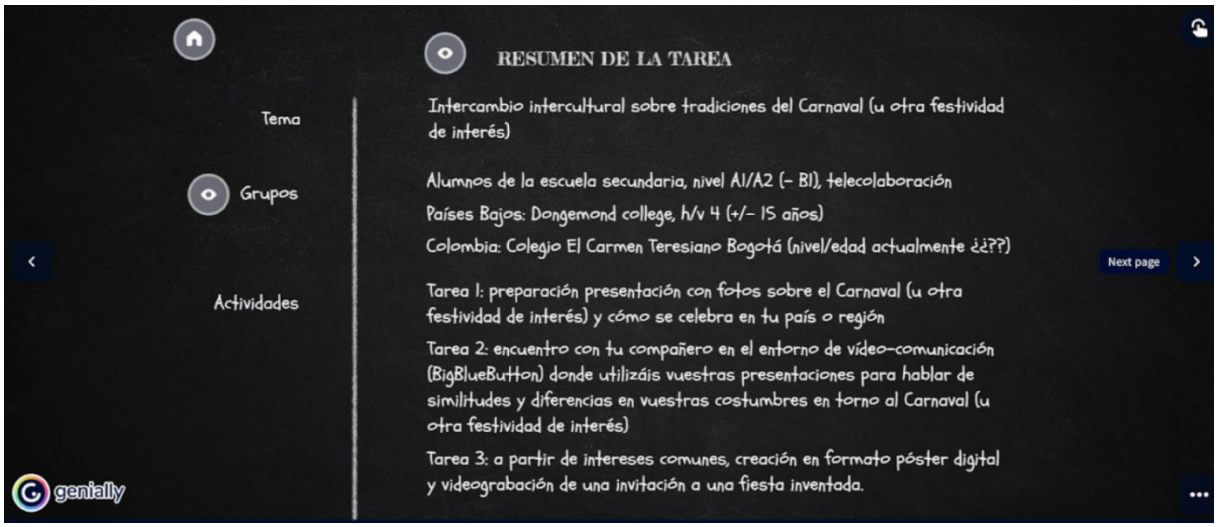


Figure 3. Sample task production 2.

In this VE task designed by trainee teachers from the Netherlands and Colombia, their students at secondary schools were prompted to engage in an intercultural exchange focusing on the traditions of Carnival (or any other festival of interest concrete learners might embrace). The task consists of three activities (see Figure 3). In the first one, students are required to prepare a presentation with photos about Carnival (or another festival) celebrated in their respective countries or regions. They need to showcase how the festival is celebrated, highlighting its customs and traditions. For the second activity, students have a video communication session where they are expected to share their presentations with their assigned partners and engage in a conversation about the similarities and differences in their customs related to Carnival or the chosen festival. In the third activity, students are tasked with creating jointly a digital poster and recording a video invitation for an imaginary party. The theme of the party should be based on common interests identified during the previous tasks. This activity encourages creativity and collaboration as students work together to design an appealing invitation that reflects elements of their cultures and shared interests.

The study investigated participants' perceptions of VE activities[1], and the results consistently demonstrated positive feedback across various aspects of the experience, as shown in Appendix C. Both Flip and the BigBlueButton (BBB) video-communication environment were considered easy to use, with mean ratings of 4.0 (SD = 1.0) and 3.9 (SD = 1.1), respectively. Additionally, participants reported high satisfaction with the sound quality in BBB, which received a mean rating of 4.1 (SD = 0.7). Regarding interaction, participants expressed enjoyment in meeting students from other countries (Mean = 4.3, SD = 0.8) and engaging in tasks with their partners (Mean = 3.7, SD = 0.9). They also felt comfortable during the sessions (Mean = 4.2, SD = 0.7) and enjoyed communicating with their international counterparts (Mean = 4.5, SD = 0.6). The VE sessions were perceived as valuable for discovering new perspectives and insights (Mean = 4.5, SD = 0.5), and fostering active participation and interaction (Mean = 4.6, SD = 0.5). The international dimension, involving interaction with students from other countries, was particularly interesting to participants (Mean = 4.4, SD = 0.8). Furthermore, participants expressed a strong desire to incorporate VE activities into their future teaching practices (Mean = 4.1, SD = 1.0).

Trainee teachers expressed their enjoyment and satisfaction with various aspects of the VE experience, including connecting with international partners, learning about different cultures, and collaborating on projects. The utilisation of new technologies and software tools was also well-received. Participants valued the exchanges with their partners, the insights into different educational systems, and the personal connections formed during the collaboration. However, trainee teachers did encounter some challenges during their participation in the VE project, such as time constraints, technical difficulties with platforms and tools, and occasional communication challenges.

Overall, the participants found the VE experience enriching, enjoyable, and informative. Their feedback highlights the benefits of VE, with one participant expressing, "*I liked meeting new people, having discussions with them, and learning about our different approaches to our future profession*."

These findings underscore the positive impact of VE activities in facilitating intercultural learning experiences, promoting engagement, and enhancing participants' awareness of the pedagogical value of VE for their future careers as language teachers. The study suggests that incorporating VE into language teaching practices can be a beneficial and rewarding approach.

4. Conclusions and perspectives

The conclusions drawn from this study serve as a critical reflection on the E-LIVE project, shedding light on both the opportunities and challenges associated with VE in language education. The E-LIVE project's primary objective is to enhance language teacher training programs and L2 teaching at primary and secondary schools through the integration of VE. While our conclusions may echo some well-established principles, they underscore the crucial role of VE in addressing specific challenges faced by language teachers. These challenges include the integration of multilingual pupils and the effective incorporation of digital tools in language education.

One key takeaway is the pivotal importance of equipping trainee teachers with the pedagogical, intercultural, and digital skills necessary to navigate the complexities of a digital world successfully. While this may align with established best practices, our project's pedagogical design showcases the innovative potential of VE. Through a variety of digital tools and platforms, trainee teachers designed tasks that promote collaboration, intercultural exchange, and authentic language use. This practical demonstration underscores the transformative possibilities of VE in language education. Moreover, the study recognizes the need for comprehensive training and support for teachers in the technical aspects of VE and the effective integration of digital tools. Clear instructions and well-organised communication channels are essential components of a successful VE experience. Additionally, the study acknowledges the time constraints that both teachers and students face, emphasizing the importance of balancing project objectives with existing workloads.

[1] The survey was validated in previous projects (TILA & TeCoLa) and adapted to the specific educational context.

Despite these challenges, it is worth noting that many trainee teachers found the VE experience enriching, enjoyable, and valuable. The connections made with international partners, exposure to different cultures, and collaborative projects were highly appreciated. These positive aspects reaffirm the potential and benefits of virtual exchanges in language education.

In essence, the E-LIVE project serves as a valuable learning experience, highlighting the evolving landscape of language education through VEs. While some findings may resonate with established principles, they reinforce the significance of VE in addressing specific challenges and fostering innovation in language teacher training. By considering the feedback and insights from trainee teachers and acknowledging both the opportunities and challenges, future projects can build upon this foundation to create even more impactful and rewarding learning experiences through VEs.

Acknowledgements

The E-LIVE project is financed by Erasmus+ KA220-SCH Cooperation partnerships in school education. We are grateful to the associate partner schools and trainee teachers who participated in the initial project actions.

References

Brammerts, H. (2002). Principes et objectifs. In Helmling, B. (Ed.) *L'apprentissage autonome des langues en tandem*. Didier.

Clavel-Arroitia, B., Wigham, C.R., Pennock-Speck, B., Jauregi-Ondarra, K. & Canto, S. Diversity in profiles and training needs of teachers engaging in virtual exchange. *Diversity in the Digital Foreign Language Classroom Conference*, March 2023, Stuttgart, Germany. pp.1 - 23. ⟨hal-04052937⟩

Directorate-General for Education, Youth, Sport and Culture (DG EAC) Digital Education unit. (2021). *Digital Education Action Plan*. https://education.ec.europa.eu/focus-topics/digital-education/action-plan

E-LIVE (2022). E-LIVE project website. https://eliveproject.eu

E-LIVE (2022b). E-LIVE YouTube Channel. https://www.youtube.com/@e-liveproject7359/playlists

Jauregi-Ondarra, K. & Canto, S. (2023). (Eds). *E-LIVE Consortium: Teacher education modules for virtual exchange.* https://eliveproject.eu

Jauregi, K. & Melchor-Couto, S. (2017). The role of coaching in teacher competence development for telecollaboration, *Alsic* [Online], 20(2), https://doi.org/10.4000/alsic.3149

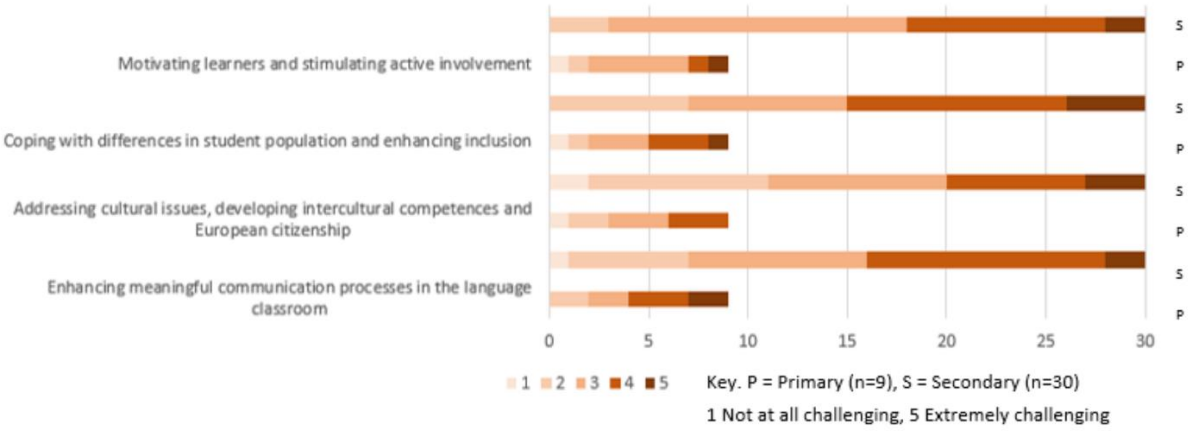
O'Dowd, R. (2018). From telecollaboration to virtual exchange: state-of-the-art and the role of UNICollaboration in moving forward. Journal of Virtual Exchange, 1, 1-23. Research-publishing.net. https://doi.org/10.14705/rpnet.2018.jve.1

Appendices

Appendix A: European-funded projects on telecollaborative practices

Project name	Educational level	Targeted languages	Resources	Website
E-LIVE Engaging Languages in Intercultural Virtual Exchange (2022-2024)	Primary, secondary, and vocational education. Teacher education institutions	English, French, German, Spanish	Teacher education modules for pre- and in-service teachers. Telecollaboration tasks for trainee teachers and school students. Case studies & participants' voices.	elive.project.eu
EVOLVE Evidence-Validated Online Learning through Virtual Exchange (2018-2020)	University	General, ELF	Monitoring study on awareness and use of VE in HEIs. Evolve report on impact of VE on teacher competences and development. Case studies research report.	evolve-erasmus.eu
EVALUATE Evaluating and Upscaling Telecollaborative Teacher Education (2017-2019)	University	General, ELF		Information available via Uni-collaboration
NIFLAR Networked Interaction for Foreign Language Acquisition and research (2009-2011)	University (Pilots at secondary schools)	Dutch, Portuguese, Russian, Spanish, Valencian	Tasks Best practices Research results	niflar.eu
TeCoLa Pedagogical differentiation through telecollaboration and gamification for intercultural and content integrated language learning (2016-2019)	Primary, secondary, and vocational education	English, French, German, Spanish	Tasks Teacher training materials Case studies Teachers' and students' experiences	tecola.eu
TILA Telecollaboration for Intercultural Language Acquisition (2013-2015)	Secondary schools	English, French, German, Spanish	Telecollaboration tasks in different languages Teacher training materials Case studies Best practices	tila.eu
Uni-Collaboration	University	General, ELF	Teacher networking	unicollaboration.org
VALIANT Virtual Innovation and Support Networks (2021-2024)	Secondary education in rural areas		Training and mentoring teachers	valiantproject.eu

Appendix B: Pedagogical Challenges in the current teaching situation



Appendix C: Trainee teachers' perceptions of VE activities (n=22) (5-point Likert-Scale: 1 Totally disagree, 5 Totally agree) (SD: Standard Deviation)

Item	Mean	SD
It was easy to use Flip.	4.0	1.0
It was easy to use the BBB environment in Moodle.	3.9	1.1
Sound was good in BBB.	4.1	0.7
I like to meet students from other countries in this environment.	4.3	0.8
I liked the tasks I carried out with my partner.	3.7	0.9
I felt comfortable during the sessions	4.2	0.7
I enjoyed communicating with a student from another country.	4.5	0.6
The VE sessions helped me discover new things about other people's views, another culture, another educational system.	4.5	0.5
I did my best to keep the interaction going during the virtual exchanges.	4.6	0.5
The international dimension (the interaction with students from another country) made the sessions more interesting.	4.4	0.8
I would like to use virtual exchange activities with my students.	4.1	1.0