PURPOSE BEFORE THE TOOL: AN AI-INTEGRATED METHODOLOGY FOR ESP LANGUAGE ACQUISITION IN HIGHER EDUCATION. AN APPLIED LINGUISTICS PERSPECTIVE

IVA BLAŽEVIĆ
FRANCESCO COSTANTINI
CRISTINA DIMULESCU
ELEONORA GALLITELLI
JELENA GUGIĆ
MOJCA KOMPARA LUKANČIČ
EMANUELA LI DESTRI
ANDREEA NECHIFOR

Transilvania
University
of Brasov

Purpose before the Tool: An AI-Integrated Methodology for ESP Language Acquisition in Higher Education. An Applied Linguistics Perspective

Iva Blažević, Francesco Costantini, Cristina Dimulescu, Eleonora Gallitelli, Jelena Gugić, Mojca Kompara Lukančič, Emanuela Li Destri, Andreea Nechifor

Transilvania University Publishing House

2026

Disclaimer

This research was supported by the Erasmus+ grant number 2024-1-RO01-KA220-HED-000248845, funded by the European Union.

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the ANPCDEFP. Neither the European Union nor the ANPCDEFP can be held responsible for them.

Acknowledgement

The authors gratefully acknowledge the support of the Erasmus+ project *'Empowering Specialised Language Acquisition with Integrated AI' (ELITE-AI)* (grant number: 2024-1-RO01-KA220-HED-000248845), within the framework of which the work supporting this monograph was carried out. The project brings together collaborative efforts to advance the integration of artificial intelligence (AI) into higher education (HE), with a particular focus on specialised language learning and English for Specific Purposes (ESP). Its activities encompass the review and synthesis of recent research, the exploration of emerging pedagogical frameworks, the integration of AI tools into language educational practice, and the identification of effective approaches informed by empirical observations. This monograph builds on these collective efforts and contributes to the ongoing discussion of the opportunities and challenges associated with the use of AI in language acquisition.

Table of Contents

TABLE OF CONTENTS **5**

INTRODUCTION **9**

CHAPTER ONE. BASIC CONCEPTS RELATED TO LANGUAGE ACQUISITION (LA), LANGUAGE FOR SPECIFIC PURPOSE (LSP), ENGLISH FOR SPECIFIC PURPOSES (ESP), AND ENGLISH AS A LINGUA FRANCA .. **20**

1.1 An Overview of Language Concepts 20
1.2 Second Language Acquisition (SLA) 21
1.3 Language for Specific Purposes (LSP) 23
1.4 English for Specific Purposes (ESP) 24
1.4.1 English for Academic Purposes (EAP) 27
1.4.2 English for Occupational Purposes (EOP) 27
1.4.3 English for Science and Technology (EST) 28
1.4.4 English for Vocational Purposes (EVP) 28
1.4.5 English for Professional Purposes (EPP) 28
1.5 English as Lingua Franca 29

CHAPTER TWO. AI IN LANGUAGE ACQUISITION – AN OVERVIEW **32**

2.1 Introduction 32
2.2 AI in Language Teaching and Learning 33
2.3 AI and Language Skills 37
2.3.1 Listening and Speaking 37
2.3.2 Reading and Writing 39
2.4 Ethical Considerations 41
2.5 Conclusion 42

CHAPTER THREE. AN AI TOOLS REPOSITORY AND AN INTERACTIVE LESSON PLAN: A METHODOLOGICAL ACCOUNT OF A PRACTICAL MAP IN AI ASSISTED ESP CLASS **44**

3.1 Introduction 44
3.2 Theoretical Background: Taxonomies, Tools, and Tutors in CALL Research 47
3.3 From Literature Review to Repository: Building the Inventory 49
3.4 From Inventory to Instrument: The Uniform Lesson Plan Template 52
3.5 From Template to Interface: The Interactive Map in Practice 56

3.6 A Grid with More than Two Axes: What the Map Actually Classifies 59
3.7 The Trial's Findings against the Wider Experimental Record 61
3.8 What the Architecture Leaves Open 64
3.9 Conclusion: Between the Inventory and the Trial 66

CHAPTER FOUR. OBSERVATION AND SELF-ASSESSMENT IN AI-SUPPORTED LANGUAGE EDUCATION: THE ELITE-AI FRAMEWORK **67**

4.1 Introduction 67
4.2 Observation and Self-Assessment as Tools for Understanding Learning Processes 69
4.2.1 Observation as a Research Tool in Educational Contexts *69*
4.2.2 Self-Assessment as a Tool for Learning and Educational Research 72
4.3 ELITE-AI Observation and Self-Assessment Sheets 74
4.3.1 Teacher Observation Sheet 77
4.3.2 Student Self-Assessment Sheet for In-Class Activities 81
4.3.3 Student Self-Assessment Sheet for At-Home Activities 83
4.4 Conclusions: methodological contribution, critical discussion and limitations 86

CHAPTER FIVE. RESEARCH AND METHODOLOGICAL FRAMEWORK FOR INTEGRATING ARTIFICIAL INTELLIGENCE TOOLS INTO ENGLISH FOR SPECIFIC PURPOSES TEACHING **90**

5.1 Introduction 90
5.2 Research Protocol 92
5.3 Overview of Research Studies Conducted as Part of the ELITE-AI Project 97
5.4 Conclusion 108

CHAPTER SIX. ANALYSIS OF AI-INTEGRATED LANGUAGE TEACHING PRACTICES: DEVELOPED CONCEPTS AND GOOD PRACTICE MODELS .. **111**

6.1 An Overview 111
6.2 Roles of AI in Language Teaching 112
6.2.1 AI as a Drafting and Revision Tool 113
6.2.2 AI as a Task Generator and Learning Resource Designer 114
6.2.3 AI as a Mediation Tool between Learners and Authentic Input 114
6.2.4 AI as a Trigger for Critical Evaluation and Metacognitive Development . 116

6.2.5 AI as Support for Genre-based and Professional Language Learning 117
6.2.6 Dynamic and Context-dependent Roles of AI 118
6.3 Typology of AI-Integrated Activities 118
6.3.1 AI-Assisted Writing with Iterative Revision 119
6.3.2 AI-Generated Practice and Student-Designed Tasks 120
6.3.3 AI-Supported Role-Play and Simulation Tasks 121
6.3.4 AI-Mediated Reading and Vocabulary Development 122
6.3.5 AI vs Authoritative Sources Comparison Tasks 123
6.3.6 Genre Transformation and Audience Adaptation Tasks 124
6.4 Cross-Context Patterns in Activity Implementation 125
6.4.1 Variation in Effectiveness across Task Types 126
6.4.2 Development of Critical AI Literacy 127
6.4.3. The Role of Prompt Design and Task Constraints 127
6.4.4. Evolution of the Teacher's Role 128
6.4.5. Conditions for Effective AI Integration 130
6.5 Implications for Future Practice 130
6.5.1. From Tool Adoption to Pedagogical Integration 130
6.5.2. Importance of Task Design 131
6.5.3. Developing Responsible AI Use 131
6.5.4. Supporting Genre Awareness and Professional Communication 132
6.5.5. Teacher Training 132
6.5.6. Enhancing Learner Autonomy 133
6.6. Conclusion 133

CHAPTER SEVEN. APPLIED GUIDANCE FOR RESEARCHERS, STUDENTS, TEACHERS, EVALUATORS, AND POLICY MAKERS. MANUALS FOR STAKEHOLDERS **135**

7.1 Introduction 135
7.2 The Pedagogical-Function Principle across the Five Manuals 137
7.3 Evidence-Based Instrument Design: Observation Sheets and the WP3 Architecture 140
7.3.1 The Observation Sheet as Research Instrument 140
7.3.2 Lesson Plans as Research and Teaching Outputs 141
7.4 Empirical Findings across Language Skills 143
7.4.1 Writing and Professional Communication 144
7.4.2 Vocabulary and Domain Terminology 145
7.4.3 Listening and Speaking 146
7.5 Autonomy, Roles and the Progression of Responsibility 146
7.5.1 Teacher Roles across a Semester 146

7.5.2 Learner Autonomy Profiles 147
7.5.3 Feedback and Critical Evaluation 148
7.5.4 Connecting Grammatical and Pragmatic Competence 149
7.6 Ethics, Disclosure and Institutional Governance 150
7.6.1 Disclosure Norms in the Classroom 150
7.6.2 Publisher Requirements 151
7.6.3 Policy Maker Perspectives and Institutional Precedents 151
7.6.4 The Equity Dimension of Digitalisation 153
7.7 Conclusion: Toward Coherent Guidance across Audiences 154

CONCLUSIONS **156**

The Decision before the Tool 156
Perception, Use and Achievement 158
Skills, Domains and the Limits of Fluency 159
Teachers and Learners 161
Critical AI Literacy in the ESP Curriculum 163
Limitations 165
Directions for Research and Practice 166
A Closing Perspective 168

REFERENCES **169**

ANNEXES **195**

Annex 1 – Teacher Observation Sheet 195
Annex 2 - Student Self-Assessment Sheet_ In-Class Activities 201
Annex 3 - Student Self-Assessment Sheet_ At-Home Activities 207

Introduction

'What is language?' is a question that was subject to evolutionary transformation through time and space. Indeed, language is the medium that enables all human existence and social interaction among different cultures and beings. Language is more than just a system of communication, it is a bridge that enables people to interact, participate in communities, express identity, develop and exchange knowledge and engage in professional, academic and intercultural contexts. As argued by Grosman (2008), language can be viewed as 'our home', a home where we live throughout our lives, but it is also viewed as a space that is constantly subject to changes, due to personal and societal experience, interaction and cultural encounters. In the present interconnected and technologically driven times, the meaning of language has evolved even more, especially in the frame of globalization, mobility, and digitalization and it continues to shape the way people acquire and use a language.

This monograph represents a vehicle that drives the reader through the study of language acquisition, which is the central area of research and educational practice, and spans through seven chapters. Each member of the implementation team of the ELITE-AI project[1] contributed to the complexity of the shared vision, with Mojca Kompara Lukančič coordinating the theoretical groundwork on language acquisition, Second

[1] Empowering Specialised Language Acquisition with Integrated AI (ELITE-AI) – Erasmus+ Project 2024-1-RO01-KA220-HED-000248845, running between 2024-2026, with a consortium consisting of the following partner universities: Transilvania University of Brasov (Romania – leader and coordinator), University of Maribor (Slovenia), University of Udine (Italy) and Juraj Dobrila University of Pula (Croatia). More details can be accessed at: https://elite-ai.unitbv.ro/

Language Acquisition (SLA) and Language for Specific Purposes (LSP) (Chapter One) and Jelena Gugić overseeing the review of research on the role of artificial intelligence (AI) in language acquisition and learning (Chapter Two). Andreea Nechifor and Cristina Dimulescu supervised two chapters, one presenting the AI Tools Repository and the Interactive Lesson Plan as the project's response to the problem of choosing tools for a stated pedagogical purpose (Chapter Three), and the other amassing the guidance the partnership produced for researchers, teachers, students, quality evaluators and policy makers (Chapter Seven). The research instruments and procedures were orchestrated by two teams, with Francesco Costantini and Eleonora Gallitelli directing, respectively, sections 1-2 and 3-4 of the chapter on observation and self-assessment as complementary methods (Chapter Four), and Jelena Gugić and Iva Blažević leading the presentation of the shared research protocol and the review of the studies published on its basis (Chapter Five). Emanuela Li Destri, in turn, steered the analysis of the AI-integrated lessons taught at the partner universities and their organisation into models that other teachers can adapt (Chapter Six). Afterwards, all the authors commented on one another's chapters, reviewing, adjusting and proofreading them and agreeing on the final form of the monograph through a collective and collaborative effort, as the complexity of the research required.

Thus, the opening chapter outlines the basic notions of language acquisition, investigating how individuals acquire a language beyond their mother tongue, mediated by concepts such as Second Language Acquisition (SLA), also examining the cognitive, social and environmental factors that influence language learning. The analysis then contrasts language acquisition through formal instruction with acquisition through

natural exposure, before turning to contemporary language education, which goes beyond general communicative competence to the specialised linguistic knowledge that professional and academic settings require. In view of the growing demands of globalization and international cooperation the focus is on the umbrella term Language for Specific Purposes (LSP) and the sub-term English for Specific Purpose (ESP) that have emerged as essential components in contemporary language acquisition. The section then focuses on the communicative needs of learners in specific professional, academic, scientific and occupational settings, where language serves as a communicative tool in achieving objectives. The presence of English is particularly noteworthy, as is has established itself as the dominant Lingua Franca of international communication and serves as the primary language in science, technology, business, higher education, diplomacy, and professional practices that span over all continents. This spread of the English language in a professional setting has created new challenges and opportunities for language instructors, who in designing learning systems in addition to linguistic accuracy also address topics related to genre awareness, intercultural communication and communication skills related to the working environment.

The second chapter is devoted to a highly relevant topic, i.e., the role of artificial intelligence (AI) in language acquisition and learning, by addressing the subject of societal changes and technological turns. The widespread availability of AI tools has transformed language teaching and learning, especially in the field of English as a Foreign Language (EFL). Thus, the development of AI-based technologies, such as intelligent tutoring systems, chatbots, speech-recognition applications, automated

writing evaluation tools, and generative AI models, are discussed with the emphasis on their capacity to personalise the learning process, adapt to learners' needs and support language acquisition through interactive and data-driven environments. The impact of AI on the development of language skills is addressed through the analyses that demonstrate how AI tools can enhance listening and speaking through speech-recognition technologies and conversational platforms, while reading and writing benefit from personalised support, immediate feedback, and automated error correction. As presented in the research, AI-assisted learning often leads to increased engagement, motivation, confidence and language proficiency. The chapter highlights the importance of a balanced approach and argues that excessive reliance on AI-generated content may reduce independent thinking, creativity and authenticity in writing skills. The section addresses, at the same time, the ethical and pedagogical implications of integrating AI into language learning by further discussing concerns related to privacy, data protection, transparency, academic integrity and highlights the importance of addressing such issues alongside the advantages of AI. As supported by the entire monograph, AI is not a replacement for language instructors, but it is a powerful complementary tool that can enrich teaching and learning.

The third chapter moves from the literature to the design implementation of the ELITE-AI project and focuses on answering the question regarding how teachers can be helped to choose an AI tool for a stated pedagogical reason instead of for its availability or popularity, starting from a weakness that research on computer-assisted language learning (CALL) has documented for some time. Stockwell (2007) found that published CALL studies often named the technology they had used without explaining why

it suited the skill in question, and Golonka et al. (2014) concluded that general claims about the effectiveness of technology in language teaching mean little unless the type of technology, the skill and the conditions of use are specified. The chapter relies on two project outputs as a single response to that weakness. One is the *AI Tools Repository* (ELITE-AI Consortium, 2025c), which catalogues 100 tools, each tagged for format, for the skills it supports, for its grammar or vocabulary focus where relevant, and for the stage of instruction (planning, teaching, learning or assessing) at which it is useful. The other one is the *Interactive Lesson Plan* (ELITE-AI Consortium, 2025d), which presents 20 lesson plans built from that inventory on a common template: a 100-minute session for 35 second-year students at B2 or C1 level, moving through lead-in, teaching, practice, feedback and homework, with each partner anchoring its plans in the field its own students study, namely academic mobility in Brașov, criminal justice and security in Maribor, tourism in Udine and education in Pula. The authors ground this architecture on Levy's (1997) distinction between 'the computer as tutor' and 'the computer as tool', Chapelle's (2001) criteria for evaluating CALL tasks, Mishra and Koehler's (2006) model of technological pedagogical content knowledge (TPCK) and Nunan's (2004) account of task-based sequencing. With the help of the cognitive-load research of Sweller (1988) and Mayer and Moreno (2003), they are also interested in understanding how many tools a learner can reasonably be expected to operate with within a ten-minute activity. The chapter closes by setting the project's trial results against recent meta-analyses and by naming what the design leaves open, including the fact that tools are renamed or disappear far faster than the instructional functions they serve.

The fourth chapter shifts the attention from what was planned to how the effects of the research vision could be studied. In this respect, starting from the known fact that tests show what learners achieve, but reveal little about how learners interact with an AI system, which strategies they adopt or how teachers fit the technology into a lesson, the consortium decided to combine two methods that capture different aspects of the same process. Observation gives access to behaviour as it unfolds in the classroom, and the chapter draws on Wragg (1994), Robson (2002) and Cohen et al. (2018) to explain what structured observation can and cannot record. Self-assessment reaches what observation cannot, namely the perceptions and strategies behind a learner's behaviour, and is discussed through the work of Boud (1995), Butler and Winne (1995), Nicol and Macfarlane-Dick (2006) and Andrade (2019), the last of whom argues that the main purpose of self-assessment is the feedback learners generate for themselves. The chapter then presents the three instruments developed for the project, the *Teacher observation sheet*, the *Student self-assessment sheet for in-class activities* and the *Student self-assessment sheet for at-home activities*, and explains what each was designed to capture, including patterns of interaction and engagement, the quality of students' prompts, their awareness of *hallucinations* and their readiness to check what an AI tool tells them. The authors are straightforward about the limits of these instruments, considering that predefined categories simplify classroom life, teachers who observe their own lessons may see what they expect to see, and self-reports are exposed to recall and effects expected to be recorded from a social desirable perspective; moreover, none of the sheets files the prompts and outputs that actually passed between a student and a machine. The templates of all three instruments are reproduced in the Annexes.

The research protocol shared by the four partner universities is set out in the fifth chapter, where the studies the partnership has published on its basis are also reviewed. After ethical approval and informed consent were obtained from the host universities' bodies and from the students, participants were divided into control and experimental groups and were submitted to taking an initial test. What followed were twelve teaching hours in which AI tools were introduced into the activities of the experimental group only while at the end of the trial all subjects sat the same final test. In the experimental group, the teacher completed an observation sheet and the students a self-assessment sheet after every class, reflecting on their activities at the same time as mirroring their experiences, and at the end of the cycle both groups reported on their use of AI at home through a dedicated observation sheet. The chapter insists on a distinction that recurs throughout the rest of the monograph: whether students found a tool useful, how they actually used it and how they performed on a test are separate kinds of evidence, and none of them can stand in for another, an aspect which the studies reviewed illustrate from several sides. Gugić and Blažević (2026) found no statistically significant difference in final achievement once initial scores were taken into account, while noting that their experimental group differed from the control group in learner autonomy as well as in the use of AI. Nechifor (2026b) discovered that most students judged AI-supported activities helpful, although many used AI chiefly because they had been asked to and did not always check what it produced. Smajla and Kompara Lukančič (2025) and Smajla and Podovšovnik (2026) documented how students use AI in their LSP learning and what concerns them about it. Kompara Lukančič and Smajla (2025) and Kompara Lukančič (2026a) applied the *Krajšavar* algorithm to English

and Italian specialised texts, Nechifor, Turculeț and Vasilescu (2025) evaluated some 200 learning scenarios generated by five platforms against pragmatic, linguistic, cultural, discourse and pedagogical criteria, and Costantini, Gallitelli and Li Destri (2026) situated the project within recent research on AI in education and ESP.

With the sixth chapter, the emphasis is laid on the lessons themselves, highlighting the AI-integrated activities that the partner institutions taught in tourism, education, criminal justice and academic mobility, in sessions of 45 to 100 minutes to learners ranging from first-year undergraduates to postgraduates. Thus, the section looks for patterns that recur across contexts and organises them into models that other teachers can adapt for their own AI-assisted teaching process. First, what is identified are the roles AI came to play in practice: a tool for drafting and revision, a generator of exercises and materials, a mediator between learners and authentic texts, a trigger for critical evaluation and a support for learning professional genres, roles that usually overlap within a single activity. Six types of activities are then described as transferable models, among them writing organised around cycles of revision, practice tasks that students design by prompting an AI tool, role-plays and simulations of professional encounters, and comparisons between AI output and authoritative sources such as official documents. The cross-context analysis that follows shows that students engaged with AI more critically in communicative, context-rich tasks than in controlled grammar exercises, and that they began to question its output only when a task required them to. It also shows that the quality of AI-supported learning depended far more on how tasks and prompts were designed than on which tool was used. Consequently, the teacher who emerges from this analysis carries more responsibility than

before, both as the designer of tasks in which AI has a defined function and as the person who knows the field well enough to show students where the machine is wrong.

The seventh chapter draws together the applied guidance the partnership produced for its different audiences, namely researchers, teachers and students, evaluators of quality and policy makers (ELITE-AI Consortium, 2026a, 2026b, 2026c, 2026d, 2026e). Instead of summarising the manuals one by one, it follows the themes they share and the evidence on which they rest: the 1,870 responses gathered through the observation and self-assessment sheets over two semesters, the presentations given at the workshop *Integrating AI Support in ESP Classes* on 7 May 2026, and the *Stakeholder Workshop for Policy Makers* held on 17 June 2026. A single basic principle, that a tool's pedagogical function must be settled before the tool is adopted, governs all of the manuals, and the chapter traces it through the design of the observation sheets and lesson plans, the findings for each language skill, the gradual transfer of responsibility from teacher to learner over a semester, and the questions of disclosure, publishers' requirements, institutional policy and equitable access raised by the trial and by the policy makers' workshop. Writing produced the most detailed evidence and specialised vocabulary the most consistent gains, while listening and speaking yielded the least conclusive results. The hesitation students showed when asked whether they had used AI ethically in their assignments suggests that disclosure is a matter individual learners cannot be expected to settle alone. The chapter ends by observing that the competence most often found lacking, the ability to evaluate AI output critically and to disclose its use, responds to direct instruction but is not yet taught in any consistent way.

The concluding chapter brings all these perspectives together, by emphasising what the evidence gathered across the project allows the authors to claim about AI in ESP teaching and where that evidence stops, and it suggests directions for research and practice beyond the end of the project.

Read as a whole, the monograph is less concerned with whether AI belongs in the ESP classroom than with its implementation and how decisions about its use should be made. Each chapter, approaching the subject from its own angle, finds that the value of an AI tool lies in the decisions surrounding it: which communicative need it serves, at what point in a lesson it is introduced, under whose guidance it is used and how its output is checked. This is the needs-based principle that Hutchinson and Waters (1987) placed at the centre of ESP, in contemporary times, applied to a new kind of resource, and it explains why the authors return so often to the role of the teacher, which needs to be redefined. A general-purpose model can produce fluent English on almost any professional subject, yet the chapters show repeatedly that fluency is no guarantee of terminological or pragmatic accuracy, and that the evaluator best placed to notice the difference is a teacher who knows both the language and the field.

Looking beyond the end of the project, several developments appear both desirable and within reach. Accordingly, catalogues of AI tools could be organised around the functions the tools perform, so that a resource compiled today remains useful after the platforms it names have been replaced. Then, the ability to verify an AI system's output against authoritative sources and to declare its use could be taught consciously, as part of the skills training curriculum, just like genre, register and

terminology already are, which represents a direction that European policy now supports: when Article 4 of the *AI Act on AI Literacy* was replaced in July 2026, the accompanying preamble depicted the development of AI literacy, beginning in education and training and continuing throughout life, as crucial (Regulation (EU) 2026/1744, preamble, para. 8). Moreover, domain-specific tools designed for authentic professional texts, such as the *Krajšavar* algorithm, which has already been extended from Slovene to English and Italian (Kompara Lukančič & Smajla, 2025; Kompara Lukančič, 2026a), suggest how specialised language teaching might come to depend less on general-purpose models, and how the approach developed here for English could be carried over to LSP courses in the other languages of the consortium, and beyond.

Thus, research will need longer interventions and designs that record what learners actually ask of an AI tool and what they do with its answers. Through all these changes the purpose of ESP remains the one described in Chapter 1, which is to prepare people to communicate with other people in their professional lives. AI helps learners reach that goal, with a legitimate place in the ESP classroom, and the chapters that follow show how that place can be defined and tested in practice.

CHAPTER ONE

Basic Concepts related to Language Acquisition (LA), Language for Specific Purpose (LSP), English for Specific Purposes (ESP), and English as a Lingua Franca

1.1 An Overview of Language Concepts

The present chapter discusses the basic concepts related to language, focusing primarily on language acquisition and, within this framework, on Language for Specific Purposes (LSP) and English for Specific Purposes (ESP). In addition, it highlights the importance of the English language and its role as a lingua franca, starting with the introduction of a quotation of particular significance, especially in the current context of widespread changes across multiple domains, including the integration of artificial intelligence (AI) into many aspects of everyday life, as well as the ongoing debate and concern regarding the potential replacement of humans by AI.

> Language is our home, the home where we have to feel good, for that reason it is worth to adapt it according to our needs, as we are going to live in it till our end, for that reason it is worth to discover and respect the homes of others in all their beauty and power (Grosman, 2008).

Apart from being our home (Grosman, 2008), language is often described as a fundamental asset that distinguishes humans from other animals. As argued by Chomsky (cf. McGilvray, 2014), humans possess language, whereas other creatures have communication systems rather than languages. Repka (2020) emphasises the importance of acknowledging the first landmark definition of language proposed by Ferdinand de Saussure (1916), which is widely adopted and which also marked the beginning of

modern linguistics. Saussure's approach has influenced numerous linguistic theories and has endured, albeit with modifications and adaptations, up to the present day. Language is indeed 'a complicated business,' as stated by Downes (1998), referring to the way in which a collection of words with individual meanings may change when combined into sentences. At the same time, it is important to recognise that a single word can be used in multiple ways depending on context.

Language can also be defined within a more conceptual framework, beginning with the classification proposed by Gass and Selinker (2008), which includes the notion of the native language (NL), defined as the first language acquired by a child (Klein, 1986), also referred to as the primary language, mother tongue, or L1. This is followed by the concept of the target language (TL), that is, the language being learned, and by Second Language Acquisition (SLA), which refers to the process of learning another language after the native language. In this context, the second language is denoted as L2 and generally refers to any language acquired subsequent to L1 (Klein 1986).

1.2 Second Language Acquisition (SLA)

From a linguistic perspective, language acquisition is most commonly discussed in terms of Second Language Acquisition (SLA), which refers to the acquisition of a second language (Doughty & Long, 2003; Gass & Selinker, 2008; Ortega, 2014; VanPatten, 2003; VanPatten & Benati, 2015; VanPatten & Williams, 2015; White, 2003). Saville-Troike and Barto (2017) define SLA as encompassing both individuals and groups who learn a language other than the one(s) acquired in childhood. It thus refers not

only to the process of learning a second language (L2), but also potentially a third, fourth, or subsequent language. Such acquisition typically occurs either in classroom settings or through direct exposure to the target language (Gass & Selinker, 2008).

Gass et al. (1988) emphasise the importance of distinguishing between foreign language learning and SLA. In foreign language learning, the target language is not used outside the classroom (VanPatten & Benati, 2015), as in the case of Italian speakers learning English in Italy. By contrast, SLA refers to the process of learning a non-native language within a natural language environment, such as Italian speakers learning English in the United States or the United Kingdom, where exposure to native speakers is readily available and learning is not necessarily confined to formal instruction.

According to Klein (1986), SLA, whether it occurs through formal instruction or everyday communication, follows a set of underlying principles related to language processing. These include, among other factors, learner motivation and the ways in which linguistic input is made available to the learner. The aim of SLA research is to identify these principles in order to support more effective language teaching and learning.

Furthermore, SLA places significant emphasis on the nature and function of human language. Gass and Selinker (2008) argue that SLA involves the creation of new linguistic systems shaped by exposure to the constraints of the second language. In this regard, researchers examine both what has and has not been acquired, while also addressing questions of native-like proficiency in multiple languages and why such proficiency is achieved by only some learners.

1.3 Language for Specific Purposes (LSP)

Considering the concept of language acquisition, a distinction must be further made between two domains: the acquisition of general language knowledge and that of field-specific language knowledge. With regard to the latter, the teaching and learning of Language for Specific Purposes (LSP) began to develop significantly after World War II (Starfield, 2013).

Language for Specific Purposes (LSP), as defined by García Laborda (2011) and Basturkmen and Elder (2004), refers to the teaching and learning of a language as a second or foreign language, with instruction oriented toward the communicative needs of learners in specific professional or occupational contexts. The teaching and learning process is therefore shaped by the communicative requirements of individuals engaged in particular work-related settings and communicative events (Kompara Lukančič, 2023).

As Tudor (1997) argues, LSP differs from Language for General Purposes (LGP) in that it is needs-based and typically constrained by limited time frames, particularly in the case of adult learners (Robinson, 1991). Over recent decades, factors such as globalization, evolving labour market demands, the internationalization of education, immigration, multiculturalism, technological advancement, and academic social responsibility have contributed to the increasing demand for and development of LSP (Uber Grosse & Voght, 2012).

At the university level, globalization and internationalization have further intensified the need for LSP courses and programs worldwide. Proficiency in multiple languages and familiarity with different cultures can enhance

employability and provide a significant competitive advantage in the job market (Grosse, 2004). Moreover, immigration and multiculturalism have heightened awareness of the need for LSP programs, as well as the demand for individuals with professional language competence capable of serving diverse domestic clients (Uber Grosse & Voght, 2012).

The concept of LSP encompasses the teaching and learning of various foreign languages (e.g., English, Italian, Spanish, French, German), with a focus on domain-specific knowledge, such as legal or medical language. Among these, English holds a particularly prominent position, giving rise to the widely recognised field of English for Specific Purposes (ESP).

1.4 English for Specific Purposes (ESP)

English for Specific Purposes (ESP) was first introduced as an area of technical text analysis in the 1960s (González Ramírez, 2015; Rahman, 2015), although its early origins can be traced back to the end of the Second World War (Hutchinson & Waters, 1987). From a historical perspective, Johns (2013) provides an overview of its development by distinguishing four main periods: the Early Years (1962–1981), the Recent Past (1981–1990), the Modern Era (1990–2011), and the Future (2011 onwards).

According to Johns (2013), in the period following World War II, ESP research primarily focused on the development of English for Science and Technology (EST) within academic contexts. In the subsequent period (1981–1990), attention shifted toward teacher training, vocational ESP, interlanguage, and the preparation of international teaching assistants.

During the Modern Era (1990–2011), research increasingly emphasised genre analysis and corpus studies.

As González Ramírez (2015) notes, the evolution of ESP has been interpreted by some scholars as an approach and by others as a distinct discipline; however, it has consistently remained oriented toward addressing learners' specific language needs. Rahman (2015) argues that the emergence of ESP was driven by several factors, including the end of the Second World War, the rapid development of science and technology, the growth in international student mobility, and the increasing use of English in fields such as science, technology, and business. From the 1960s onwards, broader economic changes further contributed to the rise of ESP as a discipline (González Ramírez, 2015).

During this period, the ESP teaching movement developed largely in response to learners with clearly defined purposes for acquiring English, particularly to meet professional requirements (González Ramírez, 2015). ESP is inherently learner-centred, with its primary aim being the fulfilment of the specific needs of target learners in order to meet professional demands (González Ramírez, 2015; Hutchinson & Waters, 1987). As Widodo (2016) argues, the objective of ESP instruction is to enable specialist learners to function effectively in professional contexts where English serves as a medium of communication.

In this sense, ESP focuses on learners' needs and involves three central components: the learners themselves, the target language, and the learning context (Howatt, 1984; Hutchinson & Waters, 1987). It is therefore

associated with specialised domains (Paltridge & Starfield, 2013) and aims to address the specific requirements of English learners (Tsao, 2011).

However, several challenges in ESP teaching have been identified. Fitria (2020) highlights issues such as the low quality of lectures and teaching materials, insufficient teacher qualifications, inadequate teaching methods, and the lack of a strong theoretical framework. Robinson (1991) emphasises that, as a goal-oriented approach, ESP requires course design based on a thorough analysis of students' specific needs. Additional challenges relate to learners themselves, including demographic factors, varying levels of English proficiency, language differences, limited vocabulary, and insufficient skills in using reference tools such as dictionaries (Fitria, 2020). Furthermore, a lack of appropriate teaching materials remains a significant issue, as also noted by Kompara Lukančič (2023) in the field of criminal justice and security.

Overall, ESP encompasses education, training, and practice focused on specialised areas of learners' interests, supported by appropriate methodologies and an emphasis on the language, skills, discourse, and genres relevant to specific professional fields (Dudley-Evans & St John, 1998; Strevens, 1980).

As defined by Strevens (1980), ESP serves as an umbrella term encompassing several subfields, including English for Academic Purposes (EAP), English for Occupational Purposes (EOP), and English for Science and Technology (EST). Dudley-Evans (1997) further identifies English for Vocational Purposes (EVP) and English for Professional Purposes (EPP) as additional subcategories. In the following subsections, a brief outline of

these subfields is provided, in a more detailed approach offered by Kompara Lukančič (2023).

1.4.1 English for Academic Purposes (EAP)

English for Academic Purposes (EAP) focuses on language education tailored to specific linguistic needs, discourse practices, and communicative skills within academic settings. Its primary aim is to address subject-specific requirements and to enhance learners' competence in studying and conducting research in English (Hyland, 2006, 2014). EAP emphasises the composition and structure of both written and spoken discourse, making these features explicit and accessible to students. Durrant (2009) highlights the growing interest in developing lists of generic 'academic' vocabulary as well as 'sub-technical' terms (Yang, 1986), which are common across disciplines but often present challenges for learners.

1.4.2 English for Occupational Purposes (EOP)

English for Occupational Purposes (EOP) focuses on enhancing learners' English language skills in relation to their professional work (Kim, 2008). As Sudipa (2020) notes, its primary aim is to meet workplace needs; therefore, instructional materials are carefully designed to align with learners' specific professional contexts and requirements. EOP primarily targets adult learners (Liu & Alley, 2019; Woods, 2018; Shiukashvili, 2020; Sudipa et al., 2020) and emphasises the acquisition of linguistic knowledge relevant to particular occupations. Sudipa et al. (2020) highlight that EOP includes work-related content from fields such as English for Tourism, English for Business, English for Science and Technology, and English for

Medicine, among others, with a strong focus on developing the linguistic competencies required in a globalised workplace (Xie, 2016).

1.4.3 English for Science and Technology (EST)

English for Science and Technology (EST) emerged over 50 years ago, at a time when science and technology became central areas of research within the field of ESP (Parkinson, 2013). EST primarily targets individuals working in scientific and technological fields and focuses on the English language as used in scientific publications, research papers, textbooks, technical reports, and academic lectures. According to Díaz Santos (2000), topics related to science and technology, such as medical issues or computer security, have traditionally been employed in EST instruction to help students acquire specialised language knowledge in these domains.

1.4.4 English for Vocational Purposes (EVP)

English for Vocational Purposes (EVP) is designed for students seeking to learn English in relation to their specific vocations (Widodo, 2016) and to enable effective vocational communication within a given professional domain (Grucza, 2009). EVP focuses on the language used for training in specific trades and occupations. For example, an English course for trainee doctors prepares students for doctor–patient interactions (Dudley-Evans & St John, 2009), emphasising the acquisition of vocabulary and technical terminology relevant to the profession.

1.4.5 English for Professional Purposes (EPP)

English for Professional Purposes (EPP) addresses the language needs of future professionals in the workplace (Ypsilandis & Kantaridou, 2007). EPP typically targets intermediate and advanced adult learners, as well as

older students, and focuses on the acquisition of specialised vocabulary and professional jargon, along with the development of effective communication skills for both workplace tasks and everyday professional interactions (Gavrilova & Trostina, 2014).

1.5 English as Lingua Franca

The term *lingua franca* is frequently used in reference to the English language and its global role. According to Samarin (1987), a language can be considered a *lingua franca* only based on its function. Mallette (2014) explains that the original concept of a *lingua franca* referred to a pidginised Romance language used in medieval and early modern Mediterranean port cities and aboard ships by speakers of mutually unintelligible native languages. This *lingua franca* incorporated elements of Arabic, Turkish, Greek, and various Romance languages or dialects commonly spoken by ordinary people. As Mallette (2014) notes, it was primarily a spoken medium of communication rather than a written language, which makes it challenging to reconstruct its grammar and vocabulary or to trace its precise historical and geographical development.

A more recent understanding of the term *lingua franca* associates it primarily with the English language. Historically, the end of the Second World War marked a significant turning point, as English emerged as the new global *lingua franca*. Several factors contributed to this development, including its widespread use and its status as an international language employed in cross-national communication and the business world (Teodorescu, 2010). English became an international language (Dudley-Evans, 2001) and, in the post-war period, also functioned as a neutral

medium as well as the principal language of science, technology, and business (Hutchinson & Waters, 1987). Consequently, there was a growing global tendency to learn English. Today, English is often described as a 'fluid, flexible, contingent, hybrid, and deeply intercultural' language (Dewey, 2007; Jenkins et al., 2011, p. 284).

As argued by Fenilli et al. (2025), research on English as a *lingua franca* (ELF) has expanded significantly in recent years, encompassing a wide range of topics. Studies have focused on the conceptualization of ELF, exploring both theoretical perspectives in global and local contexts, as well as the ideological implications of English use. A particular emphasis has been placed on pragmatics, linguistic analysis, and pragmatic strategies for intercultural communication and translation. Fenilli et al. (2025) also highlight multilingualism and translanguaging as central areas of investigation, alongside growing interest in ELF pedagogy.

ELF has been examined by numerous scholars (Fenilli et al., 2025; O'Regan, 2014; Schmitz, 2017; Berns, 2009; Kachru, 1996; Seidlhofer, 2001, 2011; Jenkins, 2007, 2009, 2013; Mauranen & Ranta, 2009; Jenkins et al., 2011), who have explored diverse subfields, including ELF awareness, Global Englishes, Critical Language Pedagogy, teacher education and attitudes, and the implementation of ELF-informed practices in specific contexts. Key movements and frameworks within the field include the 'English as a *lingua franca* movement' (Elder & Davies, 2006; Holliday, 2008; Berns, 2009) and the 'Global Englishes paradigm' (O'Regan, 2014). Additionally, ELF research recognises and accepts the varieties of English used by non-native speakers as legitimate forms of the language (Jenkins et al., 2011, pp. 283–284).

Fenilli et al. (2025) highlight recent studies that aim to (re)conceptualise English as a *lingua franca* (ELF) pedagogy, providing directions and practical examples of classroom activities. The authors also identify emerging research areas, including Study Abroad (SA), intercultural communication, English as a Medium of Instruction (EMI), and English as the Academic Lingua Franca (ELFA), as well as its applications in business (BELF), medicine (MELF), and the tourism sector. ELF is globally pervasive (Jenkins et al., 2011), present across all societies and multiple domains, including education, business, trade, diplomacy, politics, finance, culture, and even conflict (O'Regan, 2014). As O'Regan (2014) asserts, English functions as the *de facto* first language of globalization and serves as the *lingua franca* for the world.

The development and global spread of science and technology, along with internationalization and globalization, have elevated English to its current status as the *lingua franca* of international communication. The influence of English is also evident in specialised languages, professional jargon, and commercial and business contexts. As Račková and Hošková-Mayerová (2013) argue, language knowledge and the ability to communicate in a foreign language cannot be limited to grammar and general vocabulary alone; they extend to the capacity to use the language effectively in diverse contexts, where specialised language skills and professional communication are essential for international interaction.

Consequently, it can be said that English is the *lingua franca*, English is a shared language, English is a vehicle for more effective human-to-human interactions and better understanding of other languages and cultures (Moeller & Catalano, 2015).

CHAPTER TWO

AI in Language Acquisition – An Overview

2.1 Introduction

In a society where every piece of necessary information is immediately at hand and where students are not used to great efforts in accomplishing tasks, it was only a matter of time before artificial intelligence (AI) tools would enter the world of education. As Sal Khan, the founder of Khan Academy, put it in a TED talk podcast: 'We're scared of AI, but we're also embracing AI. We think kids will just give up on learning, but we also think AI will be the great equaliser.' (Khan, 2023). This duality somehow represents the fear of those working in the education system, language instructors in particular, that the great effort they are putting into teaching their students a language properly will be undermined by AI tools which will, in a way, serve as a modern 'cheat sheet' helping students overcome difficulties in an easy way. As provided in the *Guidelines on the Ethical Use of Artificial Intelligence and Data in Teaching and Learning for Educators,*

> AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. (European Commission, 2026: 8).

As Rahman (2009) put it, AI can filter knowledge and do independent tasks like computing or student search. On the other hand, as AI has entered all

the spheres of life, and thus the field of education as well, professionals have started to embrace its full potential and to treat it as a teammate, fellow worker or, as usually defined, assistant. And it was not only that professionals have started to feel at ease, but learners as well. The language instructors' fear that they could be replaced by AI tools, as irrational as it may seem, is something that has started to coexist alongside the growth in the number and quality of these tools. However, 'an artificial intelligence-based English class model should instead be developed in conjunction with the teaching and learning of English' (Ghafar et al., 2023: 19), which could help AI not to be seen as a threat.

When it comes to teaching and learning English as a foreign language, the positive impact of AI tools use is extremely prominent (Fattah et al., 2023; Vadivel et al., 2023; Aijun, 2024, Biju et al., 2024). The complex nature of teaching a foreign language encompasses the language skills which need to be mastered (namely, listening, speaking, reading and writing), while to reach them it is necessary to acquire grammatical knowledge and expand vocabulary. Therefore, the creators of AI tools for language acquisition must cater for these specific needs and have thus devised technologies such as intelligent tutoring systems (ITSs), chatbots, automated writing evaluation (AWE), natural language processing (NLP), speech recognition, and generative AI models (Ahmed et al., 2023; Benabdallah, 2023; Ghafar et al., 2023).

2.2 AI in Language Teaching and Learning

As widely known, foreign language classrooms represent a challenge for both teachers and learners (Patekar, 2017; Debreli & Ishanova, 2019;

Kankam, 2025). The nature of a foreign language classroom is characterised not only by the need to acquire the language, but also by the fact that the knowledge (language) to be acquired has to be taught and learned through the language itself. Hence the complexity of the process, considering that the medium and the object of teaching and learning are the same. The role of language instructors is therefore more demanding, as they explain language by using language (Riordan, 2018). In this sense, it can be expected that sometimes they will lack resources or ideas which could make their work easier and more interesting to language learners, while the common heterogeneity of students in language classrooms makes the role of the language instructor even more complex (Lim, 2024; Sahlender et al., 2024). When defining new generations as 'tech-savvy'[2] learners' (Hashim, 2018, as cited in Tessensohn et al., 2025: 59), the authors implied the need for teachers to adapt their teaching methods to address the needs of such learners, and the easiest way of doing that is by the introduction of new technologies in the classroom. The introduction of technology in line with the students' interests boosts motivation for learning and engagement (Pazilah et al., 2019; Pinter, 2019) and leads to higher levels of language proficiency (Martin et al., 2019). The numerous AI tools used in educational contexts adjust content based on the learner's level (Shah et al., 2025; Khosravi et al., 2022)

Today's AI used in language classes is based on scientifically trained language models (Lee et al., 2026). According to Chang et al. (2024), language models (LMs) are computational models that have the capability

[2] The online Cambridge Dictionary defines the adjective 'tech-savvy' as knowing a lot about modern technology, especially computers (retrieved on 23 July 2026 from: https://dictionary.cambridge.org/dictionary/english/tech-savvy)

to predict word sequences and generate human language if given input. With the passing of time and further technological advancements, early language models (LMs), which were primarily used for language modelling and text generation, gave dominance to large language models (LLMs) used for task solving (Zhao et al., 2023). Authors (Zaho et al., 2023; Chang et al, 2024; Naveed et al, 2024) define large language models (LLMs) as huge, advanced language models with excellent learning capabilities. Due to the aforementioned characteristics of LLMs, when pre-trained, LLMs become the base for tools facilitating language acquisition, such as the already mentioned intelligent tutoring systems (ITSs), chatbots, automated writing evaluation (AWE), natural language processing (NLP), speech recognition, and generative AI models.

When it comes to language learning, what students need the most is to get the opportunity to understand their mistakes and to learn how to correct them. Bearing this in mind, intelligent tutoring systems (ITSs), such as *e-Teacher, Duolingo,* and *REALP*, were created to provide a personalised approach to learning by acting as a teacher/expert who adapts to and manages learning according to the specific needs of each learner (Tafazoli, Gómez María & Huertas Abril, 2019; Tessensohn et al., 2025). Considering ITSs as the next step of computer-assisted instruction, Burns and Capps (1988) state that they have to pass three intelligence tests: (i) the content has to be understood by the computer system which should be able to solve problems related to it; (ii) the system must estimate the approximate level of knowledge of the students who will use it; (iii) the computer must approach different levels of performance and reduce differences among them. Corbett et al. (1997) provided eight principles and a basic one (Principle 0) governing ITSs. Namely, they stated that an ITS should enable

the student to work to the successful conclusion of problem solving (Principle 0), and then they followed with the other eight principles of what ITSs should do:

1. Represent student competence as a production set.
2. Communicate the goal structure underlying the problem solving.
3. Provide instruction in the problem-solving context.
4. Promote an abstract understanding of the problem-solving knowledge.
5. Minimise working memory load.
6. Provide immediate feedback on errors.
7. Adjust the grain size of instruction with learning.
8. Facilitate successive approximations to the target skill.

The next step in the advancement of AI technologies is the development of generative AI models such as ChatGPT where GPT stands for Generative Pre-trained Transformer. According to the European Commission's *Guidelines* 'generative AI is a specific category of AI tools and systems designed not only to analyse data but to adapt and generate content such as text, images, code, or music' (2026: 8). This is a type of AI that can generate human-like text and creative multimodal content, but its most important feature remains the possibility to collect and organise data from different sources (Dasborough, 2023, as cited in Fui-Hoon Nah et al., 2023). In language teaching, generative AI tools have proven to be extremely useful in content creation, tasks and tests design, and problem solving. As far as

learners are concerned, AI tools are used not only to find information, but very often to fulfil tasks and find solutions to problems. As these tools support the mastering of all language skills by providing tailored learning environments (Ahmed et al., 2023), they have been increasingly used in all educational fields.

2.3 AI and Language Skills

Foreign language learning has recently become an area increasingly influenced by the resources offered by AI tools. Both learners and language instructors have the opportunity to efficiently engage with personalised activities offered by AI tools. Their ability to adapt learning materials to the needs of all the participants in the learning process, create interactive learning activities, and provide immediate feedback are features that have become extremely attractive to the modern educational system. The fast-evolving nature of AI and the numerous advantages it offers make it suitable for the acquisition of knowledge in all areas. In this sense, the following chapters explore how AI tools help and support the development of the four language skills and emphasise the possible benefits and drawbacks, as well as their practical application in the pedagogical environment.

2.3.1 Listening and Speaking

When it comes to language skills, listening, as the basic skill necessary for an individual to develop all the other skills (i.e., reading, and writing), and speaking, as the main skill to be acquired and the main objective of foreign language learning, are the spotlight for numerous AI applications. Given

that mastering the listening and speaking skills is a challenging task for learners of a foreign language, 'the integration of artificial intelligence (AI) has emerged as a promising tool to address the challenges faced by ESL students' (Vincent et al., 2025: 1833). AI significantly enhances English language instruction, particularly the speaking and listening skills, although with a stronger impact on speaking (Jantakoon et al., 2025). After carrying out a study where they compared the level to which listening and speaking skills were mastered by a group of students who used AI tools and a group of students who learned in a classical way, Vincent et al. (2025) found that AI-driven platforms significantly enhanced performance in ESL learning environments. On the other hand, the group studying without the use of AI tools made improvements, but they were modest and homogeneous, without the individualised growth observed in the AI Tool Group. AI-driven platforms are particularly effective in enabling the development of complex skills like critical thinking, creativity and interactive communication necessary for language improvement at a higher level. Thus, besides enhancing listening skills, AI-driven speech recognition technology reduces listening anxiety among EFL learners (Xiao, 2025), AI tools having also shown great motivational potential (Sahito et al., 2025) because they offer diverse activities, independent and flexible practice, and reduce the shortcomings of traditional teaching methods which suggests that they should be integrated in traditional curricula by finding a balance between the two approaches (Sahito et al., 2025). As similarly suggested by Jantakoon et al. (2025), AI tools should be considered supplements, not replacements for teacher-led instruction. They thus advocate for 'a blended learning model that pairs self-directed use of technology with structured guidance from educators' (Ibid., 2025: 16).

2.3.2 Reading and Writing

Reading is a receptive language skill for which a lot of mechanical training is needed, but to be able to read and understand the content is still something which a lot of students struggle with, while teachers have a hard time finding a way to make students read since the commonly known expression 'less is more' does not apply to reading practice. As Tessensohn et al. (2025) concluded, AI tools can be used for a variety of purposes, e.g. reading comprehension, reading accuracy, reading assessment, critical thinking skills, pre-reading skills and enhancement of reading among special needs students, and for all age groups. The introduction of AI tools such as *ReadTheory* and *ChatGPT* leads to a more personalised learning experience compared to traditional methods in the sense that they provide better clarification of texts, build confidence, and support the development of consistent reading habits (Shah et al., 2025). Like with other language skills, students tend to over-rely on the AI reading tools asking them to create summaries or paraphrase texts instead of reading the texts themselves.

Considering the amount of work posed before a teacher who needs to correct, assess and grade students' written assignments, it does not come as a surprise that technological advancements in the form of ITS applications have been designed to help teachers in the correction process thus reducing the amount of time necessary to perform it. In a study conducted by Parra & Calero (2019) the authors found that the students' writing performance increased after using automated writing evaluation (AWE) tools (specifically *Grammarly* and *Grammark*). Immediate feedback offered by the tools resulted in fewer errors by the students (Son et al., 2025;

Nizzolino, 2024; Almutairi, Alfaifi & Saleem, 2026) and writing accuracy was also improved. In the end, it is important to mention that the students' attitude towards the use of such tools was positive. However, as Ranalli, Link & Chukarev-Hudilainen (2017) put it, AWE should only be a complement to language instructor feedback in writing classes, regardless of the fact that certain students describe their writing practice as teacher-centred, meaning that more attention is given to what kind of errors the teacher is going to find than to how the student is going to react to corrections (El Ebyary & Windeatt, 2010). Amaria (2025) has not only studied grammatical accuracy, but also structural coherence, originality, and critical thinking. In addition, they focused on how students made use of AI tools in the various stages of their writing process. The author concluded that using AI writing tools improves writing features such as grammar, vocabulary, and sentence structure which makes their writing more aligned with academic standards. Students using AI writing tools (such as *Grammarly, Jenni AI, ChatGPT, Gemini, Copy AI,* and *Jasper*) frequently reported immediate feedback and automated correction functions as modalities from which they had mostly benefitted from. However, originality and critical thinking lacked this kind of improvement due to limited pre-writing reflexion, brainstorming and planning, as well as to AI tool over-reliance. The mentioned over-reliance on AI tools diminishes independent writing development where students reduce the time spent on engaging with their written work deeply (Abbas et al., 2024; Jia & Tu, 2024; Zhai et al. 2024), consequently diminishing their personal writing style and creativity (Budiyono et al., 2025).

2.4 Ethical Considerations

The introduction of technological innovation has marked the advancement of all industrial fields throughout human history. Moreover, thanks to the accessibility of big data, and advancements in computer-based learning (especially deep learning), AI systems have started to grow fast. However, after AI entered the education system, numerous authors have started to reconsider the role of ethics, and this is particularly prominent in higher-education circles. According to Siau and Wang 'human rights, such as privacy, freedom of association, freedom of speech, right to work, nondiscrimination, and access to public services, should always be put in the first place' (Siau & Wang, 2020: 77). The right to privacy is especially challenged with the introduction of numerous AI tools which, to get fully accessed, require certain private data, such as e-mail addresses and professional affiliations, to be disclosed. According to certain authors (Ghosh, 2025; Xu et al., 2006), there is a high risk of data offered by AI being accessed and collected in an unauthorised manner, thus leading to an even more serious risk of privacy breach. Fairness in AI is also an important concept which ensures that users benefit from it in an equal manner, preventing bias and discrimination, and retaining autonomy in decision-making, while trust in AI is at the core (Ghosh, 2025). Research has also emphasised the importance of creating and designing good policies and applying all the necessary measures in protecting users (Ananny & Crawford, 2018).

The European Commission (2026) has identified five key considerations that justify the ethical use of AI in the educational process: human dignity, fairness, trust/trustworthiness, academic integrity, and justified choice.

The aim of language instructors should be to find a balance between using AI tools and retaining and maintaining the care, accuracy and thoroughness necessary in academic work. Critical evaluation of AI output remains at the core of academic activities, but at the same time original ideas and authorship need to be respected, all with the aim of not only reaching high academic achievements, but developing critical thought and academic integrity (Amaria, 2025).

2.5 Conclusion

Present-day education has been facing a huge challenge posed by the introduction of AI in the education system. Although AI technologies offer numerous benefits for both teachers and learners, for instance, assisting teachers in the creation of innovative and engaging lessons and helping them in the assessment of students' work, personalisation of materials, activities and tasks to cater for different students' needs, enhancement of motivation and engagement, immediate feedback, they also address significant pedagogical and ethical issues. The academic environment requires careful critical thinking and reflexion, and the tendency of education participants, especially learners, to rely on AI tools more than necessary emphasises the need for their cautious implementation. Policy makers, on the other hand, need to impose regulations linked to respecting data privacy, fairness and academic integrity.

Finally, it is necessary to understand that AI will not replace language instructors, but it should certainly be viewed as an addition to their professional work. One of the best solutions to this fear is to find a balance between traditional instruction and AI integration in the education process.

The human side of instructors such as empathy and critical evaluation, combined with the resources offered by AI will create a teaching-learning environment which will offer new, engaging and motivating activities accompanied by the always necessary and unavoidable knowledge, but also warmth and care given by human experts in the field.

CHAPTER THREE

An AI Tools Repository and an Interactive Lesson Plan: A Methodological Account of a Practical Map in AI Assisted ESP Class

3.1 Introduction

Any endeavour, any new approach and any project that sets out to catalogue educational technology inherits a difficulty the field has documented for at least two decades. When Stockwell (2007) collected every empirical article on computer-assisted language learning (CALL) published in four leading journals between 2001 and 2005, he found that researchers frequently reported which technology they had used without stating, in equally explicit terms, why that technology rather than another suited the language skill or objective in question. As a result, technology choice, in his sample, tracked convenience and availability at least as often as it pursued an articulated pedagogical rationale. Golonka et al. (2014), reviewing more than 350 studies for the University of Maryland's Centre for Advanced Study of Language, reached a related conclusion from a different angle: evidence for the effectiveness of language-learning technology and digital language acquisition varies so widely by technology type, and is documented with such uneven methodological rigour across that literature, that a general claim about 'technology' in language teaching is close to meaningless without specifying which technology, for which skill, under which conditions, and, more than that, for what purpose. Both findings describe the same fundamental gap, between the proliferation of tools available to a language teacher and the comparatively thin evidence

connecting any one of them to a defined instructional purpose, a conclusion that this monograph will consider as a grounding orientation principle, based on Khan & Mishra's (2024) observation (in ELITE-AI Consortium, 2025a: 78) that AI tools must be "guided by pedagogical principles", which in the more local terms of the ELITE-AI project (ELITE-AI Consortium, 2025a) is expressed as: a tool's pedagogical function must be established, at a defined stage and for a defined skill, before the tool is adopted.

Subsequently, an *AI Tools Repository* (ELITE-AI Consortium, 2025c) and an *Interactive Lesson Plan* (ELITE-AI Consortium, 2025d), projected together as a practical map, according to a coherent and comprehensive vision, were built as one answer to that gap: 100 tools, none entered into the database without being tagged for the skill, the stage and, where relevant, the grammar or vocabulary function it served, and 20 lesson plans built from that inventory to show that the tags can translate into an actual 100-minute classroom sequence. Chapter 2 traced the literature review that the repository relied on, and other chapters below will exemplify good practices from the lesson plans created to illustrate what the *Manual of Good Practices for Researchers* actually lists as Orientation Principle 1 (ELITE-AI Consortium, 2026a: 4): "The pedagogical function precedes the tool. Tools are chosen because they perform a specific instructional function at a specific lesson stage. They are not studied because they are available or popular." However, this chapter is interested in examining directly the design logic connecting the two: how a database built around Stockwell's and Golonka et al.'s concern was actually assembled, tagged and translated into a template a teacher could open and use as a source of inspiration and a starting point for further practice, and how that design logic compares with the wider methodological literature on classifying and

evaluating language-learning technology, a literature that reaches back both to Levy's (1997) foundational distinction between 'the computer as tutor' and 'the computer as tool' and to Mishra and Koehler's (2006) account of the knowledge a teacher needs to integrate in any technology that they choose to use, especially if there is a specific content area targeted at all.

Thus, this chapter addresses that comparison directly and treats the two outputs not as separate deliverables to be described in turn but as a single design problem solved in two stages and worth reading against the field's own accumulated experience. Consequently, Section 3.2 situates the repository and the lesson plan within the CALL literature on tool taxonomies and evaluation criteria, drawing on Levy (1997), Stockwell (2007), Hubbard (2006) and Chapelle (2001), as well as on Nechifor and Dimulescu's (2024) more recent account of how ESP can be accommodated to CALL; Section 3.3 describes how the repository's 100 entries were assembled and tagged, read against the literature's own evaluation criteria; Section 3.4 turns to the uniform template through which entries drawn from the repository were built into 20 lesson plans, based on Nunan's (2004) account of task-based design and on Mishra and Koehler's (2006) technological pedagogical content knowledge framework; Section 3.5 follows the template onto the live interface published at https://elite-ai.unitbv.ro/index.php/interactive-lesson-plan/ (ELITE-AI Consortium, 2025d), using the reading and writing pages as primary evidence and drawing on the cognitive-load research of Sweller (1988) and Mayer & Moreno (2003) to ask what happens when several tools are combined within a single ten-minute activity; Section 3.6 analyses the multiple classification layers the architecture encodes at once; Section 3.7 places the project trial's own findings, in dialogue with the wider experimental record

on AI and the four language skills, a record that in places supports the experiment's findings and in places complicates them; Section 3.8 considers what the architecture leaves open, and Section 3.9 closes by locating the repository and the lesson plan between the chapters that precede this one and the trial that tested them.

3.2 Theoretical Background: Taxonomies, Tools, and Tutors in CALL Research

Before turning to ELITE-AI's own repository, it is worth asking what the wider field of CALL research has already established about classifying and evaluating language-learning technology, since the repository is an instance of a genre with its own history. Levy (1997) proposed a distinction that remains one of the field's more durable reference points: a piece of software can occupy a tutor role, in which the programme itself manages the interaction, sequences the content and evaluates the learner's response, or a tool role, in which the software provides a resource, a word processor, a search engine, a corpus, that a teacher or learner uses for a purpose the software itself does not define. Hubbard (2006), writing within the same tradition, built an evaluation framework around a closely related distinction, arguing that a piece of CALL software's teaching presence, i.e., the degree to which it substitutes for rather than supports a teacher's own instructional judgment, has to be assessed before its content can be judged on its own terms. Chapelle (2001) proposed a parallel but independently derived set of five criteria for evaluating a CALL task's appropriateness: language learning potential, meaning focus, authenticity, positive impact and practicality, criteria broad enough to be applied both to a piece of

software, to a teacher-planned activity built around that software, and to a learner's actual performance while using it. Developing on Chapelle's vision, Nechifor and Dimulescu (2024) built their own mobile application, as part of the LanGuide Erasmus+[3] interdisciplinary research team, catering for the exact needs of several ESP fields and also integrating cultural elements of the target language in an original manner, thus addressing the need identified in the educational digital landscape under the name of what Nechifor (2025a) calls 'digital pragmatics'.

Read against this literature, the ELITE-AI repository's own tagging scheme looks less like an administrative convenience specific to the project and more like a compressed, applied version of distinctions the field has been refining since the 1990s. The skill tags, i.e., speaking, writing, reading and listening, and the process-stage tags, i.e., planning, teaching, learning and assessing, do for 100 entries, in a few characters each, a version of what Chapelle's five-criterion framework does in prose for a single piece of software: they force a decision about what a tool is for before it is catalogued. The *Educaplay* entry discussed further in Section 3.3, tagged across every skill and every stage, sits closer to Levy's tool function, a general-purpose resource a teacher configures for a specific purpose, while an entry such as Kompara Lukančič's *Krajšavar* (Amebis, n.d.), tagged narrowly for abbreviation recognition in written texts, sits closer to a tutor role within its own narrow domain, since the algorithm performs a specific

[3] Language Guidance Tool for Improving Language Knowledge (LANGUIDE) – Erasmus+ Project 2019-1-SI01-KA203-060377, running between 2019-2022 under the coordination of: University of Primorska (Slovenia). The mobile application created by the project partners: University of Zadar (Croatia), Transilvania University of Brasov (Romania), Mälardalen University (Sweden), University of Rijeka (Croatia), and University of Castilla-La Mancha (Spain) can be downloaded from: https://languide.si/en/the-languide-app

analytic operation the teacher does not and would otherwise have to do by hand (Kompara Lukančič, 2025, 2026a). Neither position on Levy's continuum is superior in the abstract; Hubbard's (2006) point, restated for ESP by the *Manual of Good Practices for Researchers'* Orientation Principle 1, is that a repository entry becomes usable only once a teacher knows which position a given tool occupies for the activity at hand, and knows it before the lesson rather than working it out in front of the class.

Stockwell's (2007) empirical finding, that published CALL research frequently under-specifies the pedagogical rationale for a given technology choice, and Golonka et al.'s (2014) related finding, that effectiveness evidence is too unevenly distributed across technology types to support general claims about technology as such, together describe the methodological vulnerability the ELITE-AI repository was built to address at the level of tool selection. But whether the repository actually avoids that hazard or merely displaces it to the point at which a partner fills in a lesson plan's tool field without an explicit pedagogical rationale is a question the next two sections analyse in detail, since a tag recorded in a database is a claim about pedagogical function only if the tagging process itself was disciplined, coherent and staged at the level of the entire partnership, among all the researchers involved.

3.3 From Literature Review to Repository: Building the Inventory

The *AI Tools Repository* was compiled from two sources the consortium kept deliberately separate before combining them. The first was the literature review conducted at the project's earlier stage (ELITE-AI

Consortium, 2025b), which supplied a working sense of which categories of AI application, ITSs, AWE, speech recognition, and generative language models could be considered for the practical part, as already displayed in Chapter 2, above. The second was the direct exploration: each of the four partner institutions tested more than 20 tools independently, weighing them against the practical demands of an ESP classroom rather than against their marketing claims (ELITE-AI Consortium, 2025c). The 100 entries published in the final repository are the outcome of reconciling these two sources rather than of either one alone, so that a tool's presence in the database indicates both that it appeared credible in the secondary literature and that a partner had used it either with their own students or on their own, to test it, before recommending it. Read against Chapelle's (2001) authenticity criterion, whether a task involves genuine rather than simulated language use, this compilation method did a version of that evaluative work before any lesson plan existed: a tool a partner had actually tested with ESP students had already been assessed for authenticity in Chapelle's sense, whereas a tool known only from the secondary literature had not, a distinction the repository's published tags do not currently record but that its two-stage compilation method built in as a matter of procedure rather than of documentation.

The repository organises its 100 entries alphabetically, a choice that serves search criteria rather than pedagogy: a teacher who already knows a tool's name can find it quickly, but alphabetical order carries no instructional sequence of its own. The pedagogical work is done instead by a set of tags attached to every entry, recorded independently of alphabetical position. Each tool is marked as web-based or mobile, and separately marked for the skills it targets, using letter codes for speaking, writing, reading and

listening, with a further marker distinguishing tools built for one specific skill from tools supporting a combination of several (ELITE-AI Consortium, 2025c). A parallel set of labels records which stage of the instructional process a tool suits: planning, teaching, learning or assessing.

The *Educaplay* entry shows how densely a single tool can be categorised: it is marked for all four skills and for all four process stages alike, since the platform generates word searches, matching games, gap-fill exercises and quizzes that a teacher can deploy at almost any point in a lesson. An entry such as *Connected Papers*, by contrast, is tagged narrowly for reading and writing at the teaching and learning stages, reflecting its specific function as a tool for tracing citation relationships in academic literature rather than a general-purpose classroom application. A third pattern is visible in the corpus-analytic entries, *Sketch Engine* being a representative of the type: Kilgarriff et al. (2014) describe the platform's core functions, word sketches, concordance and thesaurus generation, as tools for extracting grammatically structured collocation patterns from a corpus a user supplies, which is a categorically different operation from anything a general-purpose chatbot performs on request, and the repository's tags register that difference by restricting such entries to reading, writing and the learning and assessing stages rather than marking them, as *Educaplay* is marked, for every skill and every stage at once.

The repository's sixth stated feature is a deliberate emphasis on tools built for specialised vocabulary rather than general usage (ELITE-AI Consortium, 2025c), and the reason that emphasis matters methodologically becomes clearer against the vocabulary literature specifically. Coxhead's (2000) *Academic Word List*, still the most widely

used general resource of its kind, was built from a 3.5-million-word corpus of academic prose sampled evenly across arts, commerce, law and science, and a word qualified for inclusion only if it recurred across all four of those broad fields rather than being confined to one; the list is general precisely because field-specific terminology was excluded from it by construction.

A word central to criminal justice and security, or to the vocabulary of academic mobility and credit transfer, is for that reason largely invisible to a general academic word list, not because such words are unimportant but because Coxhead's selection criteria were designed to filter them out in favour of vocabulary that generalises across disciplines. This is the concrete reason a repository built for ESP could not simply point teachers toward an existing general-purpose vocabulary resource and needed its own identifier, distinguishing vocabulary-specific from grammar-specific tools: the needs-based account of ESP that Chapter 1 sets out, in which a course and by extension a tool introduced into it is justified by what a defined group of learners specifically needs rather than by the tool's general availability (Hutchinson & Waters, 1987), is not a preference the repository could have satisfied with a general-purpose word list even if one had been available, since the vocabulary ESP students most need to encounter is the vocabulary such lists are built to exclude.

3.4 From Inventory to Instrument: The Uniform Lesson Plan Template

A hundred tagged entries do not by themselves constitute a lesson plan, and the step from one to the other involved a second design decision distinct from the first. Nine academics from the four partner institutions, with an

average of more than 20 years of teaching experience among them, agreed on a single template that every lesson plan in the project would follow (Blažević et al., 2026). The template fixes several parameters in advance: a second-year cohort at B2 or C1 level, 35 students, a 100-minute session, and the five stages of a classical lesson continuum i.e., lead-in, teaching, practice, feedback and homework. Within that fixed frame, each partner supplied the variable elements proper to its own field: the ESP domain, the specific skill and grammar or vocabulary focus for a given activity, the type of interaction, individual, pair or group, and the AI tool drawn from the repository that best fit the aim of that particular stage.

The repository's own stage tags, planning, teaching, learning and assessing, do not map onto the lesson plan's five classroom stages in any one-to-one way, and the mismatch is worth enlarging upon. A tool tagged for planning in the repository describes who prepares material and for what general purpose; a tool placed in the lead-in stage of a lesson plan describes where, within a live 100-minute session, that preparatory function actually occurs. Turning a four-category process tag into a five-stage classroom sequence required a partner to decide, for every tool entered into a lesson plan, not only that its repository tag matched the activity but which of the five stages that activity belonged to, and why. The template supplied the frame within which this decision had to be made explicit; it did not automate the decision itself. This is why the template, rather than the repository alone, is where Orientation Principle 1 was first enforced structurally.

Mishra and Koehler's (2006) technological pedagogical content knowledge framework offers a useful way of describing what filling in a single template cell actually required of a partner. Their model holds that a teacher

integrating any technology into instruction is coordinating three kinds of knowledge at once, content knowledge of the subject being taught, pedagogical knowledge of how to teach it, and technological knowledge of what a given tool can do, and that weak integration typically results from strength in one or two of these without the third rather than from the absence of all three. Populating a single activity in the ELITE-AI template required exactly this coordination in miniature: the ESP domain and the grammar or vocabulary focus supplied the content component, the lesson stage and interaction type supplied the pedagogical component, and the tool drawn from the repository supplied the technological component, with no cell counted as complete until a partner had specified all three together. A template cell filled in with a tool but no stated skill, or a stage but no stated tool, would not satisfy the design Mishra and Koehler describe as adequate integration, and the template's own structure, by requiring all three fields before an activity could be entered, effectively prevented the two-out-of-three failure mode their framework treats as the most common cause of poor technology integration in teaching generally.

The five-stage architecture itself, lead-in, teaching, practice, feedback and homework, sits in a specific and only partial relationship to the wider literature on task design. Nunan (2004), building on Willis's (1996) earlier framework, argues that a task-based lesson is organised around a pre-task phase, a task cycle in which learners use language to accomplish a non-linguistic outcome, and a post-task phase in which form is addressed only after meaning has already been negotiated, a sequence in which grammar instruction follows communicative use rather than preceding it. The ELITE-AI template's insistence that every activity split into parallel vocabulary and grammar tracks proceeding from the same source text,

discussed further in Section 3.5, sits closer to a more conventional presentation-practice-production design than to Nunan's meaning-first sequencing, since form is addressed within the same ten-minute window as meaning rather than deferred until after a communicative outcome has been reached. This is not necessarily a shortcoming; ESP instruction of the kind Chapter 1 describes carries a heavier specialised-lexis burden than general-purpose language teaching, and Coxhead's (2000) point about the field-specific vocabulary a general course can safely defer applies with less force when the vocabulary in question is the entire reason the course exists. The observation is offered here as a design choice worth naming rather than as a departure from best practice: the template's architecture is closer to a form-and-meaning-integrated model than to a pure task-based one, and a reader coming to the lesson plans expecting Nunan's sequencing would need to recalibrate that expectation before evaluating them on their own terms.

Twenty lesson plans were produced in total, one set of five for each of the four canonical skills, spanning four ESP domains (ELITE-AI Consortium, 2025d). The domain-to-skill correspondence is visible directly in the two pages examined in Section 3.5: the reading page anchors its five activities in mobility and academic exchange, the same domain Nechifor, Turculeț and Vasilescu (2025) evaluated independently in their cross-platform assessment of mobility-scenario generation, while the writing page anchors its five activities in criminal justice and security, the domain associated with Kompara Lukančič's *Krajšavar* work at the University of Maribor (Kompara Lukančič, 2025, 2026a). Tourism, associated with the University of Udine, and education, associated with Juraj Dobrila University of Pula, supply the remaining two domains. Because each partner anchored its activities in the ESP field its own students actually study, the lesson plan

inherits the needs-based logic of Chapter 1 at the level of content, not only at the level of tool selection: a mobility-domain glossary is of limited use to a criminal justice cohort, and the design does not pretend otherwise.

One further detail follows from the template rather than from any stated design target. Five activities per skill, across four skills, produce 20 activities at every one of the five stages, for a total of 100 discrete activity slots across the full architecture (Blažević et al., 2026). This is the same order of magnitude as the 100 tools catalogued in the repository, though nothing in the project's documentation states that the two figures were meant to correspond. The lesson-plan architecture arrived, through its own interwoven combinations, at roughly the scale of the inventory it draws from, a reasonable outcome for a template built to use that inventory rather than to exhaust it with a single browse.

3.5 From Template to Interface: The Interactive Map in Practice

The paper version of the 20 lesson plans, the working sheets the nine academics actually filled in, looked like conventional lesson plans laid out on a page. The online version, published at https://elite-ai.unitbv.ro/index.php/interactive-lesson-plan/ under four skill-specific addresses for reading, writing, listening and speaking, rebuilt the same content as an interface a visitor can move through non-sequentially, selecting a skill, then a stage, then an activity, rather than reading a document from the first stage to the last (Blažević et al., 2026). This is why the deliverable works as a map with an interactive dynamics component rather than as a manuscript: a teacher preparing a listening lesson has no need to pass through the reading page first, and the interface does not require it.

The reading page, examined directly for this chapter as an example of good practice, shows how densely a single lead-in activity can combine tools drawn from the repository. Its first activity, built around Erasmus+ and student-exchange vocabulary, opens with an AI-generated word cloud through *Educaplay*, which students sort into familiar and unfamiliar terms before a short, paired discussion prepares the reading passage that follows. Each of the five activities on the page then splits into two parallel tracks proceeding from the same source text, one built around vocabulary and one around grammar, a division the authors describe in their monograph as a way of ensuring that no skill is activated in isolation from at least one other (Blažević et al., 2026). A single ten-minute activity in the teaching stage draws on *Twee* to generate the reading passage, *Text Inspector* to analyse its lexical complexity, then *Grammarly* and *PrepAI* for the grammar track running in parallel, so that four repository entries, each independently tagged, combine within one classroom activity rather than being deployed one at a time.

This pattern of assembling several independently designed tools inside a single short activity is worth examining against cognitive load research, since nothing in the repository's own tags addresses it directly. Sweller (1988) distinguishes 'intrinsic load', the difficulty inherent in the material itself, from 'unnecessary load', the difficulty imposed by how information is presented rather than by the task at hand and argues that working memory's limited capacity means superfluous load directly competes with the capacity available for learning the material itself. Mayer and Moreno (2003), extending the cognitive load theory to multimedia instruction specifically, propose a coherence principle: material that does not directly serve the learning goal should be excluded even when it seems supportive, because every additional element a learner must parse adds peripheral load

whether or not it was intended as a help. Four AI interfaces within one ten-minute activity, each with its own input conventions, output format and interaction pattern, *Educaplay*'s word cloud, *Twee*'s generated passage, *Text Inspector*'s complexity report, *Grammarly*'s correction markup, ask a B2 or C1 undergraduate to absorb four distinct operating logics inside a single window of instruction. The repository's tags establish that each of the four tools serves a distinct, independently justified pedagogical function; they do not, by themselves, establish that a single ten-minute activity is the right unit within which to ask a learner to operate all four, and the trial data, discussed further in Section 3.7, were not designed to measure cognitive load directly. The point raised here is accordingly a question the architecture leaves open rather than a criticism its own evidence base can currently settle.

The interactive lesson plan dedicated to writing confirms the pattern that will be reiterated by the current monograph's subsequent chapters: its five activities, built around correspondence, dictionary compilation, abbreviations, forensic linguistics and legal terminology, all sit within the criminal justice and security domain, and the abbreviations activity illustrates Orientation Principle 1 in operation, tracing *Mentimeter* through *Krajšavar, DeepL, Grammarly* and *SketchEngine* across the five stages of a single 100-minute session. Analysing the page dedicated to writing directly shows one further detail: the same five topics recur unchanged across all five stages of the writing page, so that a student who begins the lead-in stage on legal terminology completes homework on legal terminology rather than moving to a different topic halfway through. This is consistent with the template's own requirement that each of the 20 lesson plans follow a single ESP scenario from lead-in to homework, rather than treating the five stages as interchangeable slots that could be filled from any topic at hand.

3.6 A Grid with More than Two Axes: What the Map Actually Classifies

The repository and the lesson plan can generally be described as organising AI tools by skill and by stage, and that description is accurate as far as it goes, but it understates how many classification layers the architecture carries at once. A single entry in the repository is tagged simultaneously for format, web or mobile; for skill, one or several of the four; for grammar or vocabulary specificity, where applicable; and for process stage, planning, teaching, learning or assessing. A single activity within a lesson plan adds two further layers on top of these: the classroom stage it belongs to, one of the five in the lesson continuum, and the ESP domain the whole lesson plan is anchored in. No one of these layers, taken alone, organises the material usefully, thus, a skill tag without a stage tells a teacher what a tool does but not when to use it, and a stage without a domain tells a teacher when to use a tool but not for whom.

Mishra and Koehler's (2006) 'technological pedagogical content knowledge framework' (TPCK), already introduced in Section 3.4 to describe how a single template cell was filled in, applies equally to the finished architecture read as a whole. Their model holds that no one of content knowledge, pedagogical knowledge and technological knowledge is sufficient on its own, and that the three must be coordinated by the person doing the teaching rather than supplied ready-made by a resource. The repository's tag structure can accordingly be read as an attempt to pre-compute the TPCK intersection for a teacher before the teacher ever reaches the classroom: the skill and grammar or vocabulary tags encode content, the process-stage and lesson-stage tags encode pedagogy, and the tool identity itself encodes technology, so that a completed lesson-plan cell

is a worked example of ‘technological pedagogical content knowledge’ rather than evidence that the teacher who eventually opens that lesson plan possesses the underlying judgment independently. This distinction matters because Mishra and Koehler (2006) present TPCK as a property of a teacher’s understanding, not of a resource external to the teacher. However, even if a tagged database can model the intersection of the three knowledge types, it cannot transfer the judgment behind it automatically, a limitation Section 3.8 returns to.

The first monograph’s account of the reading, writing, speaking and listening pages adds a further, cross-cutting layer that runs across all four skills rather than within any one of them: content generation, exemplified by *Twee* and *ChatGPT*; assessment and feedback, exemplified by *Grammarly, PrepAI* and *Vocabulary.com*; interactive engagement, exemplified by *Educaplay* and *Mentimeter;* text analysis, exemplified by *Text Inspector* and *Revit*a; and multimedia integration, exemplified by *YouLearn.ai* (Blažević et al., 2026). This functional layer classifies what a tool technically does, independent of which skill it happens to be deployed for in a given activity, and it is worth distinguishing from a layer the following chapters develop for a different purpose, when identifying five conceptual roles AI came to occupy once the third work package’s (WP3) trial actually brought these tools in front of the students: drafting and revision support, task generation, mediation between learners and authentic input, a trigger for critical evaluation, and support for genre-based professional communication. The two five-part schemes are not the same classification restated twice: the map’s functional layer describes a tool’s built-in capability before any teacher has used it, while the trial’s role layer describes the instructional purpose a tool ends up serving once a teacher

has actually deployed it in a specific activity, and a single tool can occupy a different role in practice than its nominal function would predict.

Taken together, these layers are where Orientation Principle 1 is encoded structurally rather than merely stated as a rule. The current chapter restates the principle in prose, as something the *Manual for Researchers* instructs a teacher to observe (ELITE-AI Consortium, 2026a); the repository and the lesson plan enforce a version of the same requirement mechanically, since an entry cannot be completed without a skill tag, a stage tag and, once it reaches a lesson plan, a domain tag. A tool without an established pedagogical function had nowhere to be entered in the first place; consequently, the manual's prescription and the map's architecture converge on the same claim from two different directions: one written after the fact as guidance, the other built in advance as a constraint on what could be catalogued at all.

3.7 The Trial's Findings against the Wider Experimental Record

The architecture described in Sections 3.2 through 3.6 was built before the WP3 trial produced any evidence about how well it would work in practice, so it is worth asking how the trial's own findings sit against the wider experimental literature on AI and second-language skill development published in the same period. The comparison is not straightforward, because the two bodies of evidence differ in scale and in what they were designed to measure, but the discrepancies are informative in their own right rather than being a reason to set the comparison aside.

Chapter 7 (Section 7.4) reports that writing produced the most detailed evidence of any skill in the trial, while listening and speaking returned the most equivocal results of the four. The comparison by Gugić and Blažević (2026) at the University of Pula, which found no significant difference in final achievement on a test covering listening, reading, vocabulary and grammar, points the same way. Two recent meta-analyses reach a different balance. Wu and Li (2026), synthesising 21 studies published between 2008 and 2023, found that AI chatbots have a significant overall effect on EFL learners' language skill development, without reporting a comparable asymmetry between productive and receptive skills. Wu (2024), integrating a larger set of 49 studies comprising 79 reports, went further and reported that AI interventions were most effective for listening and speaking specifically, with writing and vocabulary showing significant but comparatively smaller benefits, which is close to the inverse of the pattern Chapter 7 reports for the ELITE-AI trial.

Goh and Aryadoust (2025), reviewing the same listening and speaking literature specifically for the *Annual Review of Applied Linguistics*, offer a plausible reconciliation rather than a simple contradiction. They note that the dialogue-based systems most often credited with listening and speaking gains were evaluated primarily on learner perception, engagement and a sense that the system felt humanlike, and that the studies they reviewed did not systematically evaluate whether AI-driven conversation practice produced measurable linguistic development beyond that engagement effect. If a meaningful share of the evidence behind Wu's (2024) listening-and-speaking finding is closer to an engagement effect than to a measured skill gain, the discrepancy the data in Section 7.4 narrows considerably, since the WP3 observation sheets were built to record engagement and

measured outcomes as separate and potentially divergent findings rather than to collapse them into a single effect size. Read in this way, the ELITE-AI trial and the wider meta-analytic literature are not straightforwardly in conflict; they may simply be answering different questions about the same two skills, one about how students feel while using a tool and one about what a 45 to 100-minute session can be shown to change.

Where the trial and the wider literature agree more closely is on writing, and the agreement is worth stating plainly rather than passing over because it is less surprising than the listening and speaking discrepancy. Both Wu and Li (2026) and Wu (2024) report significant benefits for writing alongside their headline findings for other skills, and Shi et al. (2025), in a randomised 11-week comparison of 150 students split between a *ChatGPT* group, a conventional automated writing evaluation (AWE) group and a control group, found that the *ChatGPT* group outperformed both other groups on writing performance. Their study also found something the ELITE-AI architecture does not currently anticipate, i.e., compared with the AWE group, the *ChatGPT* group's ideal L2 writing self, a measure of how students see themselves as writers in the language, was significantly lower, even though their measured writing performance was higher. This finding complicates a design habit visible throughout the reading and writing pages described in Section 3.5, where a generative tool such as *ChatGPT* or *Twee* and a conventional AWE tool such as *Grammarly* are routinely paired within the same activity as though the two functioned as interchangeable or additive feedback sources. Shi et al.'s (2025) result suggests the two tool types may pull in different directions on a dimension the repository's tags do not currently record at all, performance against self-concept, and that combining them inside a single activity, as several of the

writing page's activities do, is a design choice with an empirical cost the project's own trial data were not positioned to detect, since the WP3 observation sheets, as reported in the chapters below, tracked motivation and language learning as broad categories rather than isolating a construct as specific as ideal L2 writing self.

3.8 What the Architecture Leaves Open

Three limitations follow from how the repository and the lesson plan were built, and none of them is hidden, as the consortium's own documentation flags at least the first directly. A footnote to the repository's Introduction, reproduced in the first monograph (Blažević et al., 2026), warns that some of the 100 tools, compiled in 2025, had already been replaced, rebranded or discontinued by the time of publication, a caution that applies with at least equal force to the present monograph's own account. A taxonomy organised around named tools ages faster than one organised around functions, since a platform can disappear while the instructional need it served, generating a differentiated reading passage, for example, or extracting collocations from a specialised corpus, remains constant. The functional layer described in Section 3.6 is accordingly the more durable part of the architecture, and any future revision of the repository would do well to treat the functional tags as the stable spine and the named tools as replaceable content attached to them.

A second limitation concerns the gap between the architecture's own even coverage and what the WP3 trial, read alongside the wider literature discussed in Section 3.7, found once these tools were actually used. The lesson-plan template distributes its 20 activities evenly, five per skill at

every stage, giving reading, writing, listening and speaking identical structural weight within the map, and nothing in the tag structure anticipates that this even weighting would translate into uneven results, or that the direction of that unevenness would run opposite to what the broader meta-analytic literature would have predicted. Equal representation in a taxonomy does not guarantee equal pedagogical return in the classroom, and a map built before any classroom trial took place, and before Wu's (2024) and Goh and Aryadoust's (2025) reviews were available to consult, had no way of anticipating which of its four evenly weighted branches would prove hardest to observe results for once the trial actually ran, or which of the field's existing findings that result would end up sitting closest to.

A third limitation is the domain-anchoring described in Section 3.4, which is at once the architecture's strongest link to needs-based ESP theory and its narrowest constraint. Because each partner built its five activities around its own students' field, academic mobility for Brașov, criminal justice for Maribor, and tourism and education for Udine and Pula respectively, a teacher working in a fifth ESP field (medicine, engineering or maritime studies, for instance) inherits the map's tool logic and its stage architecture but not its content, and would need to repeat the domain-specific work each partner already did for its own students rather than adopting a ready-made lesson plan directly. This is not a flaw so much as the necessary consequence of building an ESP resource on needs-based principles in the first place; a genuinely domain-neutral version would have had to give up the specificity that makes it useful to any one field.

3.9 Conclusion: Between the Inventory and the Trial

This chapter has traced two design decisions rather than one: the decision that produced 100 tagged entries out of a literature review and four institutions' independent testing, and the separate decision that turned a subset of those entries into 20 lesson plans built around a single shared template. Neither decision reduces to the other, and the gap between them, the translation from a four-category process tag to a five-stage classroom sequence, is where most of the architecture's actual design work happened, even though that work leaves no single artefact of its own to point to beyond the finished lesson plans. Read against the CALL literature surveyed in Section 3.2, from Levy's (1997) tutor-tool distinction to Mishra and Koehler's (2006) account of integrated teacher knowledge, the architecture looks less like a project-specific administrative solution and more like an applied instance of problems the field has been working on for three decades, solved here under the practical constraints of a four-institution Erasmus+ partnership rather than in the abstract.

The repository and the lesson plan sit, in the sequence this monograph follows, after the literature review that supplied their initial inventory and before the manuals that later cited them as evidence and the classroom trial that tested them directly. The chapters below will show what the set of manuals generated by the consortium made of a single lesson plan once the trial had run its course, and Section 3.7 of this chapter has already put some of that trial's own findings in dialogue with the wider experimental record; the chapters that follow return to the trial in full, and to how the 20 lesson plans described here performed, skill by skill and stage by stage, once they left the interface and entered an actual 100-minute session.

CHAPTER FOUR

Observation and self-assessment in AI-supported language education: The ELITE-AI framework

4.1 Introduction

As discussed in the previous chapters, the integration of Artificial Intelligence into language education raises a range of pedagogical and methodological issues. The present chapter shifts the focus from the design and implementation of AI-assisted activities to the methods through which their educational impact can be investigated.

The ELITE-AI project was designed to explore the educational potential of Generative Artificial Intelligence in English for Specific Purposes (ESP) contexts through the implementation of a wide range of AI-assisted learning activities. Beyond introducing innovative AI-assisted instructional practices into language education, the project sought to identify effective pedagogical approaches and examine their impact on learners, teachers, and researchers. Achieving these objectives required more than measuring learning outcomes alone. While tests and performance measures can provide evidence of what learners achieve, they often offer only limited insight into how learning takes place, how students interact with AI systems, which strategies they adopt, or how teachers integrate AI into their pedagogical practices.

Against this backdrop, a key methodological question emerged: how could these new practices be systematically observed, documented, and analysed in order to provide reliable evidence of their educational value? To address

this challenge, the project adopted a multidimensional approach combining observation and self-assessment. Observation provides access to externally observable behaviours and interactions, allowing researchers to document learner behaviour, interaction patterns, levels of engagement, and classroom practices as they unfold. Self-assessment instruments, in contrast, capture learners' perceptions, motivations, strategies, and interpretations of their own learning experiences. Used in combination, these tools make it possible to triangulate data and obtain a more comprehensive understanding of educational practices, documenting how participants engage with instructional activities, learning resources, and AI-supported educational technologies.

This chapter is organised into two main sections. Section 2 provides the theoretical framework underpinning the methodologies adopted in the project, discussing observation and self-assessment as complementary methodological and pedagogical tools in educational research. Particular attention is devoted to their role in documenting learning processes, supporting self-regulated learning, and generating evidence on educational practices. Section 3 presents the instruments developed within the ELITE-AI project, including the *Teacher observation sheet*, the *Student self-assessment sheet for in-c*alysis of AI-mediated language learning practices.

The templates of the three instruments used within the ELITE-AI framework are reproduced in the Annexes: the *Teacher observation sheet,* the *Student self-assessment sheet for in-class activities,* and the *Student self-assessment sheet for at-home activities*.

4.2 Observation and Self-Assessment as Tools for Understanding Learning Processes

4.2.1 Observation as a Research Tool in Educational Contexts

Observation has long been recognised as one of the most valuable methods for investigating educational processes. Unlike assessments that focus primarily on learning outcomes, observation enables researchers to examine teaching and learning as they occur in authentic contexts. Its distinctive strength lies in the possibility of collecting 'live' data from naturally occurring situations, allowing researchers to study behaviours, interactions, and educational practices directly rather than relying exclusively on participants' retrospective accounts.

Educational processes are inherently complex, involving multiple actors, interactions, and contextual variables. For this reason, observation has occupied a central place in educational research, providing access to dimensions of classroom life that are often difficult to capture through tests, interviews, or questionnaires alone. In the specific context of classroom research, Wragg (1994) emphasises that observation provides a systematic means of studying teaching and learning as they unfold in real time, making it possible to document the multifaceted nature and richness of classroom interactions as well as the diverse purposes for which classroom behaviours and educational practices may be examined. As Robson (2002) argues, observation functions as a form of 'reality check', enabling researchers to compare what participants claim to do with what actually takes place in educational settings. Similarly, Cohen et al. (2018) note that observation allows researchers to examine everyday practices that might otherwise go unnoticed or be taken for granted by participants themselves.

Observation is particularly valuable because it allows researchers to investigate not only educational outcomes but also educational processes. Morrison (1993) identifies several dimensions that can be documented through observation, including the physical setting, the human setting, the interactional setting, and the program setting. In educational environments, this means that researchers can simultaneously examine classroom organization, learner behaviour, teacher actions, interaction patterns, instructional strategies, and the use of educational resources. Such breadth makes observation particularly well suited to studying innovative and rapidly evolving learning contexts.

The importance of observation has also been widely acknowledged in language education research. Classroom observation has been used extensively to analyse teacher talk, learner participation, interaction patterns, communicative practices, and instructional approaches. Observation-based studies have significantly contributed to understanding how language learning takes place within real classroom environments and how pedagogical decisions influence learner engagement and language development (Richards & Farrell, 2005; Spada, 2019).

From a methodological perspective, researchers generally distinguish between structured and unstructured forms of observation. Unstructured observation seeks to capture events as they naturally emerge and is often associated with exploratory and qualitative research. Structured observation, by contrast, relies on predefined categories, indicators, and recording procedures that guide data collection and facilitate systematic analysis. According to Cohen et al. (2018), structured observation is particularly useful when researchers wish to compare behaviours across

different contexts, identify recurring patterns, or monitor specific aspects of educational practice.

Structured observation is characterised by the use of observation schedules or observation sheets that define in advance what is to be observed and how observations should be recorded. This approach reduces the risk that researchers focus selectively on certain events while neglecting others: through the use of carefully designed categories, observation schedules enable the collection of comparable evidence across participants, classrooms, institutions, and educational interventions. In this respect Wragg (1994) notes that systematic classroom observation requires explicit observational frameworks that guide attention toward relevant aspects of classroom activity while ensuring consistency and comparability across observations.

The educational value of structured observation extends beyond research purposes. Borich (2016) argues that systematic observation can support reflective teaching, professional development, and evidence-based pedagogical decision-making. By documenting learner engagement, interaction patterns, and classroom dynamics, observation helps teachers move beyond intuition and anecdotal impressions, providing a more systematic basis for evaluating instructional practices. In this sense, observation serves as a tool for educational improvement and not only as a research method.

Despite its strengths, observation is not without limitations. Observers can document behaviours and interactions, but obviously they cannot directly access participants' intentions, thoughts, motivations, or perceptions. Observation is also susceptible to issues such as selective attention, observer bias, expectancy effects, and difficulties related to the

interpretation of behaviour. For this reason, Cohen et al. (2018) stress the importance of careful operationalization, observer training, and methodological triangulation to strengthen the validity and reliability of observational findings. One way of addressing these limitations is to combine observational data with learners' own accounts of their experiences, perceptions, and learning processes, thereby incorporating perspectives that remain largely inaccessible through observation alone.

4.2.2 Self-Assessment as a Tool for Learning and Educational Research

Alongside observation, self-assessment has gained increasing recognition as a valuable tool for understanding learning processes and supporting educational development. More specifically, self-assessment allows learners to reflect on their own experiences, strategies, motivations, and perceptions.

Although self-assessment has been conceptualised in different ways, a common element across definitions is its focus on learners' active evaluation of their own learning. Brown and Harris (2013) describe self-assessment as a descriptive and evaluative act carried out by students regarding their work and academic abilities, while Panadero, Brown, and Strijbos (2016) define it as a set of mechanisms through which learners assess and evaluate aspects of their learning processes and products.

An early, particularly influential contribution comes from Boud (1995, 1999), who argues that self-assessment should not be viewed merely as an assessment technique but as a learning strategy through which learners assume responsibility for monitoring and improving their own performance. This perspective shifts attention from the act of judging achievement to the broader process of developing learner autonomy and reflective capacity.

A related perspective is offered by Butler and Winne (1995), who place self-assessment within the framework of self-regulated learning: effective learning depends on learners' ability to generate internal feedback about their own progress and use that information to regulate subsequent actions. Self-assessment therefore functions as a mechanism through which learners monitor performance, identify gaps between current and desired levels of achievement, and make adjustments to their strategies.

The educational value of self-assessment is closely linked to broader theories of self-regulated learning. Zimmerman and Schunk (2011) argue that successful learners actively set goals, monitor their progress, evaluate their performance, and adjust their learning strategies accordingly. In this perspective, self-assessment becomes one of the key processes supporting learner autonomy and metacognitive development.

The relationship between self-assessment and feedback has been further developed by Nicol and Macfarlane-Dick (2006), who emphasise the importance of enabling students to become active participants in the assessment process. According to their model, self-assessment contributes to learning by helping learners understand quality criteria, monitor their progress in relation to learning goals, and make informed decisions about future learning actions. Effective feedback, therefore, is not only information transmitted by teachers but also information generated by learners themselves through reflective evaluation.

More recently, Andrade (2019) has argued that the primary purpose of self-assessment is not grading but feedback. From this perspective, self-assessment should be understood as a process through which learners generate information about their own learning and use that information to

make adjustments that improve performance and deepen understanding. Consequently, self-assessment is most effective when it serves a formative function, providing opportunities for reflection, revision, and further learning rather than merely producing a judgement of achievement. In her review of 76 empirical studies, Andrade (2019) found substantial evidence that self-assessment contributes to academic achievement, autonomy, and metacognitive development (meta-analyses conducted by Graham et al. (2015) and by Panadero et al. (2017) similarly report positive effects of self-assessment on learning performance and self-regulation).

Particularly important is the role of criteria and standards in supporting effective self-assessment. Research suggests that students benefit most when self-assessment is guided by explicit criteria, rubrics, checklists, or structured prompts. Such tools help learners compare their work against clearly defined expectations and provide a framework for identifying strengths, weaknesses, and areas for improvement. Studies by Andrade et al. (2008, 2010) as well as Panadero and Romero (2014) demonstrate that rubric-referenced self-assessment can improve both learning outcomes and learners' ability to regulate their own learning.

4.3 ELITE-AI Observation and Self-Assessment Sheets

Building on the theoretical principles outlined above, the ELITE-AI project sought to embed observation and self-assessment directly into the design and implementation of its educational activities. Rather than treating these instruments as external research tools applied after the completion of teaching and learning activities, the project adopted an integrated approach

in which observation and self-reflection became part of the pedagogical process itself. This approach was consistent with the project's broader objective of investigating not only the outcomes of AI-supported language learning, but also the processes through which learners and teachers engaged with AI tools in authentic educational settings.

To achieve this goal, a multidimensional monitoring framework was developed, combining structured teacher observation with student self-assessment activities conducted both during and beyond classroom instruction. Throughout the implementation of AI-assisted language learning tasks, observation sheets were used to document classroom interactions, learner engagement, patterns of AI use, and pedagogical practices. At the same time, self-assessment instruments encouraged students to reflect on their experiences, learning strategies, perceived progress, and attitudes towards AI-mediated learning.

The observation and self-assessment sheets were organised around teaching aims and lesson stages and were designed to capture multiple dimensions of AI-supported ESP learning, including positive and negative aspects of AI use, cultural factors, domain-specific features of the different ESP contexts, and individual learning preferences. In this way, the instruments provided a common framework for documenting first-hand experiences across the different educational settings involved in the project, while remaining sensitive to the specific characteristics of each ESP domain.

The data collected through these instruments were also intended to form a shared repository of first-hand evidence on AI-supported ESP practices across the partner institutions, providing an empirical basis for the dissemination and publication activities of the project.

The following subsections describe the three instruments that were adopted for the systematic observation and evaluation of AI-mediated language learning processes within the ELITE-AI project: the *Teacher observation sheet*, the *Student self-assessment sheet for in-class activities*, and the *Student self-assessment sheet for at-home activities*.

Within the ELITE-AI framework, these instruments were conceived as the empirical backbone of the project's practical work on AI integration in ESP. WP3 addresses a complete instructional continuum – teaching, learning, and assessment – and introduces AI-supported practices across lesson stages and across different disciplinary varieties of ESP. The observation and self-assessment sheets were therefore designed not simply to record whether AI was used, but to document how its use related to pedagogical aims, lesson phases, learner needs and formative or summative assessment contexts.

A further design principle was standardization across partner institutions. The ELITE-AI project envisaged uniform sheets capable of generating a comparable repository of first-hand observations while still retaining space for contextual specificity. The project documentation explicitly identifies positive and negative aspects, cultural dimensions, domain-specific particularities, and references to personalised learning styles as relevant elements to be captured. This combination of common categories and context-sensitive information was intended to support cross-context comparison without losing sight of the distinctive characteristics of the ESP domains represented in the project.

The sheets also have a cumulative function within the project architecture. Rather than constituting isolated classroom records, they were planned as a structured database from which recurrent patterns, challenges and effective practices could be identified and subsequently translated into research publications and a manual of good practices for teachers and students. The project set quantitative targets of 100 observation sheets for each of the teaching, learning, and assessment strands, with 160 students and eight teachers involved in each strand. These figures should be understood as implementation targets defined in the project design rather than as findings reported in the present chapter.

This positioning is methodologically important. It places the ELITE-AI instruments at the intersection of research, pedagogical reflection and project quality development: they support systematic data collection, but they also create a shared language through which instructors and learners can reflect on the affordances and constraints of AI-supported ESP practices. The framework thus treats observation and self-assessment as components of an iterative evidence cycle in which classroom experience informs analysis, and analysis can in turn inform subsequent pedagogical decisions.

4.3.1 Teacher Observation Sheet

The *Teacher observation sheet* constitutes the core observational instrument of the ELITE-AI framework. It provides a structured framework through which teachers document observable behaviours, classroom dynamics, instructional practices, and the integration of AI tools into language learning activities. By focusing on directly observable phenomena, the sheet aims to generate systematic and comparable data across institutions, learning contexts, and instructional settings.

The instrument is designed to collect both contextual and pedagogical information. In Questions 1-8, teachers record institutional affiliation, course level, disciplinary field, class duration, and class size. In addition (Questions 9, 10), the sheet documents the stage of the lesson being observed, distinguishing between lead-in, teaching, and practice phases. This information is aimed at allowing researchers to relate observed behaviours to specific pedagogical contexts and instructional objectives.

A substantial part of the observation protocol (Questions 11-22) focuses on lesson design and learning objectives. Following the structure of the lesson plans developed within the ELITE-AI project (see Chapter 3 above), the observation sheet is designed to document the pedagogical objectives underpinning each AI-supported activity. In particular, instructors are asked to record the language skills associated with the selected AI tools, identifying whether the activity aims to develop reading, writing, listening, speaking, grammar, or specialised vocabulary. By linking observed classroom practices to the learning outcomes specified in the lesson plans, the observation sheet is intended to situate AI use within a broader pedagogical framework, emphasising that technological tools are employed to support clearly defined educational goals rather than functioning as ends in themselves.

The observation sheet also documents the technological environment in which learning takes place. Teachers record the specific tools employed (Question 14), as well as the materials and resources used alongside AI (Question 15), including computers, mobile devices, interactive whiteboards, projectors, classroom management platforms, worksheets, and other instructional technologies.

Particular attention is devoted to the pedagogical purposes for which AI is employed (Questions 17-22). The instrument distinguishes among the major language domains and documents the specific functions assigned to AI within each area. For listening activities, for example, the observation sheet records uses such as audio generation, transcription, listening-comprehension support, pronunciation models, and authentic material curation. Similar categories are provided for speaking, reading, writing, grammar, and vocabulary activities, allowing observers to identify patterns in the pedagogical integration of AI across different language-learning objectives.

Another major dimension concerns student engagement (Questions 23-30). Teachers evaluate the overall impact of AI use on learning and participation, documenting whether engagement appears diminished, maintained, or enhanced. More detailed indicators focus on students' active participation (Question 25), including their willingness to initiate interaction, volunteer contributions, seek clarification, and engage with assigned tasks. The instrument also records the frequency of peer interaction and the overall degree of student involvement, distinguishing between voluntary, prompted, and passive participation (Questions 26, 27).

The sheet further examines patterns of classroom interaction (Question 28). Teachers document the relative frequency of teacher-student interaction, peer collaboration, student-AI interaction, whole-class participation, group work, and individual support. These categories are intended to make it possible to analyse how AI tools influence the social organization of classroom learning and whether they facilitate, reshape, or potentially reduce forms of interpersonal communication traditionally associated with language learning environments.

A dedicated section addresses the nature and quality of student-AI interaction (Questions 31-37). Teachers evaluate students' motivation, their level of independence in deciding when and how to use AI, and the effectiveness of their interactions with AI systems. Particular attention is given to prompting practices, communication with AI tools, and learners' ability to interpret and act upon AI-generated feedback (Questions 32, 33). These indicators provide valuable evidence regarding the development of AI literacy and strategic AI use in educational contexts. Given the growing concern over the reliability of AI-generated content, the observation sheet also includes indicators relating to the critical evaluation of outputs, as teachers are asked to document whether students verify AI-generated information, whether they show awareness of *hallucinations* and potential inaccuracies, or whether they tend to accept AI responses without questioning them. Complementing this dimension, the instrument investigates the extent to which students meaningfully integrate AI into the learning process, identifying situations in which AI is used productively, in isolation from learning goals, or with excessive dependence on AI support.

The final section (Questions 34-37) focuses on learning outcomes and classroom behaviour. Teachers evaluate observable developments in language skills, including listening, speaking, reading, writing, grammar, and specialised vocabulary. They also assess behavioural dimensions such as help-seeking behaviour, interaction with the teacher, adherence to instructions, management of technological tools, avoidance of distractions, and respect for classmates. Although these observations do not constitute formal measures of achievement, they provide valuable indicators of the ways in which AI-mediated activities influence both learning processes and classroom participation.

4.3.2 Student Self-Assessment Sheet for In-Class Activities

To complement classroom observation and teacher-reported data, the ELITE-AI framework includes a student self-assessment sheet specifically designed for in-class activities. While observation protocols document learners' externally observable behaviours, the self-assessment instrument captures students' own perceptions of classroom participation, AI use, interaction patterns, and learning experiences. The instrument thus provides an additional perspective on AI-mediated language learning, enabling the triangulation of observational evidence with learners' self-reported experiences.

The questionnaire was administered after classroom activities involving the use of AI tools. In addition to collecting background information such as institution, study level, field of study, instructor, and observation session number (Questions 1-6), the instrument focused on students' experiences during a specific learning event. This design allows researchers to link learners' perceptions directly to particular classroom activities and observation records.

One section of the questionnaire investigates students' perceptions of the effectiveness of AI-supported learning (Questions 7, 8). Learners are asked to evaluate whether AI facilitated or hindered their learning experience and whether the extent of AI use during the activity was appropriate. These items provide insight into how students perceive the pedagogical value of AI tools and the role they attribute to them within classroom learning.

A second dimension concerns classroom participation (Questions 9, 10, 11). Students are asked to report the frequency with which they engaged in

various communicative behaviours, including speaking voluntarily, initiating conversations, responding to questions, seeking assistance from the teacher, and requesting support from AI tools. The instrument also explores peer interaction, a key component of communicative language learning. Students assess both the frequency and quality of their collaboration with classmates, indicating the extent to which AI-supported activities facilitated, maintained, or reduced interaction among learners. This section is especially relevant in discussions surrounding AI integration, as concerns have frequently been raised regarding the possibility that AI tools may either enhance or replace human interaction during learning tasks.

Language use constitutes another important area of investigation (Questions 12, 13). Participants evaluate how frequently they used the target language during the activity and whether they were successful in conveying their intended meanings. They also reflect on the accuracy of their language use, particularly with respect to grammar.

The questionnaire further includes items addressing motivation (Question 14) and learner autonomy in using AI tools (Question 15). Students are invited to reflect on their level of engagement during the activity and to indicate whether their use of AI was self-initiated, teacher-directed, or deliberately avoided. This question provides insight into the extent to which learners perceive AI as an autonomous learning resource rather than merely a required component of classroom instruction.

Particular attention is devoted to the quality of students' interactions with AI systems (Question 16). Learners evaluate the effectiveness of their exchanges with AI tools, the quality of their prompts or instructions, and

their ability to understand and make use of AI-generated responses. This question addresses dimensions that are closely related to prompting competence and AI literacy, both of which have become increasingly important in contemporary educational contexts.

The instrument also investigates learners' critical engagement with AI-generated content (Questions 17, 18). Students report whether they verified AI outputs and whether such verification was motivated by awareness of AI *hallucinations* or concerns about the reliability of the information provided. By including these items, the questionnaire seeks to assess the extent to which learners adopt a critical and reflective stance towards AI-generated content rather than accepting it unquestioningly.

Finally, in Question 19, students are asked to reflect on the overall integration of AI into the learning process and to evaluate their general experience of AI-assisted classroom activities. This concluding item provides a broad indication of learners' acceptance of AI-supported learning and their perceptions of its usefulness in the language classroom.

4.3.3 Student Self-Assessment Sheet for At-Home Activities

In addition to classroom observation and the student self-assessment sheet for in-class activities, the ELITE-AI framework incorporates a student self-assessment instrument designed to document learners' experiences with AI-supported language learning beyond the classroom.

The instrument was administered at the end of the semester and focused specifically on AI-assisted at-home learning activities. It collected both contextual information and reflective evaluations regarding students' use of AI tools for language learning purposes. Participants were first asked to

provide background information, including their institution, study level, academic field, and instructor (Questions 1-7). They were then invited to report which AI tools they had used throughout the semester (Question 8), selecting from a range of commonly employed platforms such as *ChatGPT, Copilot, Gemini, Claude, Perplexity, Grammarly, DeepSeek*, and others.

The questionnaire is designed to explore the purposes for which AI tools were employed (Questions 9, 10). Students could indicate multiple uses, including brainstorming, grammar checking, style improvement, translation, specialised vocabulary development, content summarization, proofreading, quiz generation, conversation practice, and *hallucination* detection. This question was intended to capture the breadth of learners' engagement with AI and to identify the functions that students perceived as most valuable in their individual learning processes.

The self-assessment sheet also examined learners' perceptions of effectiveness (Questions 11, 12). Students were asked to evaluate whether AI had made learning easier, harder, or had produced no substantial difference. Additional questions investigated the perceived appropriateness of AI use and the frequency with which learners relied on AI for homework, general language improvement, and specialised vocabulary development, providing insight into the role assumed by AI within students' learning routines.

Another important dimension concerned learner motivation and autonomy (Questions 14, 15). Students were invited to reflect on their level of engagement during the semester and to indicate whether their use of AI stemmed from personal initiative or from explicit instructional requirements. This distinction was considered particularly relevant for understanding the extent to which AI had become integrated into learners'

self-directed learning practices rather than being perceived merely as a tool imposed by the educational context.

Given the increasing importance of AI literacy in educational settings, particular attention was devoted to the quality of students' interactions with AI systems (Question 16). Several items explored how effectively learners communicated with AI tools, how successful they considered their prompts and instructions, and how well they understood and applied AI-generated responses.

The questionnaire further addressed students' critical evaluation of AI-generated content (Question 17). Participants were asked whether they verified AI outputs and, if so, whether such verification resulted from awareness of *hallucinations* or from suspicion regarding the accuracy of specific responses.

A further set of items focused on the integration of AI into the broader learning process (Questions 18-20). Students reflected on whether AI had become part of their regular study habits, whether they perceived it as a useful resource for self-learning, and whether they were able to establish meaningful connections between AI use and their language development. Ethical aspects of AI-supported learning were also considered through questions addressing the responsible integration of AI-generated content into academic assignments.

Finally, the instrument investigated students' perceived learning gains (Questions 21-25). Learners were asked to evaluate the extent to which AI-assisted activities contributed to the acquisition of specialised vocabulary, the understanding of grammar in context, the development of pragmatic

competence, the adoption of professional writing practices, and the improvement of receptive and productive language skills.

Additional items (Questions 26, 27) explored the emotional dimension of participation in the project and changes in motivation throughout the semester.

4.4 Conclusions: methodological contribution, critical discussion and limitations

The ELITE-AI framework responds to a central methodological problem in AI-supported language education: learning with generative AI cannot be adequately understood through outcome measures alone. By combining structured teacher observation with student self-assessment in class and at home, the framework makes visible several dimensions that conventional achievement measures tend to overlook, including patterns of participation, peer and teacher interaction, prompting practices, learner autonomy, critical verification of AI outputs and the perceived integration of AI into language-learning routines. In this respect, the three instruments provide complementary perspectives on the same educational process and create opportunities for triangulation between externally observed behaviour and learners' interpretations of their own experience.

A particular strength of the framework is its attempt to balance comparability and contextual sensitivity. The use of common categories across partner institutions can support systematic analysis across ESP settings, while the attention to cultural aspects and domain-specific

particularities acknowledges that AI-mediated language learning is situated rather than uniform.

The sheets can therefore function both as research instruments and as prompts for pedagogical reflection. Their value is not limited to describing classroom events: repeated use may help teachers identify patterns in engagement, interaction, dependence on AI, and critical AI literacy, while self-assessment may encourage learners to make their strategies and decisions more explicit.

At the same time, the standardization that makes the sheets useful also creates an important limitation. Predefined categories inevitably reduce the complexity of classroom activity and may privilege behaviours anticipated by the instrument designers while overlooking unexpected forms of AI use, interaction or learning. Ratings such as engagement, effectiveness, independence, or quality of interaction require interpretation and may not have identical meanings across teachers, institutions, disciplines or lesson types. Unless observers receive shared training and inter-rater agreement is examined, comparisons across contexts should therefore be treated cautiously. The dual role of teacher and observer may introduce additional expectancy and confirmation effects, particularly when instructors are also responsible for implementing the AI-supported activity being evaluated.

The student self-assessment sheets introduce a different set of constraints. Self-reports provide access to perceptions and intentions that observation cannot capture, but they remain vulnerable to social-desirability effects, inaccurate self-estimation and differences in how learners interpret response categories. The in-class sheet reduces the temporal distance between experience and reflection, yet responses may still be influenced by

the immediate success or difficulty of a particular task. The at-home instrument, administered at the end of the semester, offers a broader view of learning routines but is more exposed to recall bias and may compress heterogeneous experiences with different AI tools into a single retrospective judgement. Perceived learning gains should consequently not be interpreted as equivalent to independently measured gains in language proficiency.

There are also limitations specific to the study of generative AI. Observation sheets and questionnaires can record whether students report checking outputs, producing effective prompts, or using AI strategically, but they do not by themselves provide direct evidence of the prompts entered, the outputs received, the revisions made, or the reasoning through which learners accepted or rejected AI suggestions. Moreover, AI systems, interfaces, and model capabilities change rapidly. Categories that are appropriate for one implementation period may require revision as new tools and interaction modes emerge. The instruments should therefore be regarded as adaptable components of a monitoring framework rather than as a fixed or technology-neutral measurement scale.

These limitations should be considered in light of the broader methodological design of the ELITE-AI project. Observation and self-assessment data were triangulated with additional evidence of language development, as the teaching experiment included a placement test at the beginning and a final test at the end of each course across the four partner universities. This combination made it possible to complement teacher observations and students' perceptions with measures of language performance, thus providing different perspectives on the educational

processes investigated. Future work could further examine the reliability and validity of the rating categories, the consistency of teacher judgements across sites, and the extent to which student and teacher perspectives converge or diverge. Such analyses could help distinguish more clearly between perceived usefulness, observable engagement, strategic AI literacy and demonstrable learning outcomes.

Beyond their research function, the observation and self-assessment sheets also had a practical and formative role within the ELITE-AI project. The systematic documentation of teachers' and students' first-hand experiences was intended to inform the development of a manual of good practices for teachers and students, thereby translating evidence gathered during implementation into pedagogical guidance for the responsible and effective integration of AI into ESP teaching and learning.

Accordingly, the contribution of the ELITE-AI observation sheets lies less in offering a definitive measure of the effectiveness of AI than in providing a systematic structure for documenting how AI becomes embedded in ESP teaching, learning, and assessing. Used critically, iteratively and in combination with other forms of evidence, the framework can support both research and reflective practice. It offers a basis for identifying effective practices and recurrent problems across contexts while preserving the need to interpret AI-supported learning as a situated, socially mediated and rapidly evolving educational phenomenon.

CHAPTER FIVE

Research and Methodological Framework for Integrating Artificial Intelligence Tools into English for Specific Purposes Teaching

5.1 Introduction

The use of Artificial Intelligence tools in language learning and teaching has attracted increasing research attention in recent years (Kohnke et al., 2023; Li et al., 2025). This development is particularly relevant to English for Specific Purposes (ESP), where language teaching is shaped by learners' specific academic and professional needs (Hutchinson & Waters, 1987). In this context, the aim is not only to develop general language competence, but also to enable students to understand and use language in situations related to their field of study and future professional practice. The use of AI in ESP teaching, learning and assessing therefore cannot be evaluated solely on the basis of whether students use a particular tool. Research into the use of AI tools in the ESP teaching process, with all three of its aspects, consequently encompasses not only their use, but also the purposes for which students use them, how they approach generated content, how they assess its reliability and appropriateness, and how such use relates to the learning process and achieved outcomes.

In studies conducted within the ELITE-AI project, AI tools were incorporated into ESP teaching, learning and assessing, while the research focus was not limited to students' attitudes towards their use. The common research protocol envisaged monitoring the teaching process throughout the instructional cycle, initial and final assessment of language knowledge,

and data collection from both teacher and student perspectives. This made it possible to address different aspects of AI tool use, ranging from student and teacher experiences to language achievement. Distinguishing between these types of data is important for their interpretation because perceived usefulness, the manner and frequency of AI tool use, and test performance represent different dimensions of the learning process. A student's assessment that a particular tool facilitated content comprehension or task completion does not necessarily imply better performance on the final test. Likewise, a test score provides information about language achievement but does not in itself explain students' classroom experiences or how they used AI tools. For this reason, data within the project research framework were collected using different instruments, each focusing on a particular aspect of the teaching process or its outcomes.

The project's methodological framework is linked to the fundamental characteristics of ESP teaching, learning and assessing. Hutchinson and Waters (1987) define ESP as an approach to language teaching in which content and methods are shaped according to learners' needs and their reasons for learning the language. Within such an approach, the use of AI tools is likewise connected to students' specific linguistic and communicative needs. Dudley-Evans and St John (1998) further emphasise the importance of specialised discourse, genre characteristics, and the requirements of a particular academic or professional community. In the project research context, this provides a basis for considering not only the linguistic accuracy of content generated by AI tools but also its terminological, discourse, and communicative appropriateness for a specific professional context.

The ELITE-AI project focuses on investigating the possibilities for using AI tools in ESP teaching, learning and assessing. The research framework does not assume that their use in itself leads to better educational outcomes than traditional teaching approaches. The focus is on how AI tools are incorporated into ESP teaching and learning and on monitoring the teaching process, the experiences of teachers and students, and the educational outcomes associated with their use. AI is therefore examined within the context of the actual teaching process, alongside teacher and student experiences and educational outcomes related to its use.

The purpose of this chapter is therefore to present and analyse the research approach and methodological framework used to investigate the application of AI tools in ESP teaching, learning and assessing within the ELITE-AI project. The focus is on the common research protocol, the organisation of research in the teaching environment, data collection instruments and procedures, and the methodological approaches represented in project studies. Rather than treating project publications as separate papers, this chapter considers them in relation to the aspects of the research problem they address, allowing the methodology to be examined through specific research procedures and findings.

5.2 Research Protocol

The common research protocol provides the basis for systematically monitoring the use of AI tools in ESP teaching, learning and assessing. Before the research began, the protocol envisaged obtaining approval from the relevant ethics committee, informing students about the purpose and

procedure of the study, and obtaining their consent to participate. This was followed by the assessment of the initial level of language competence of students in the control and experimental groups.

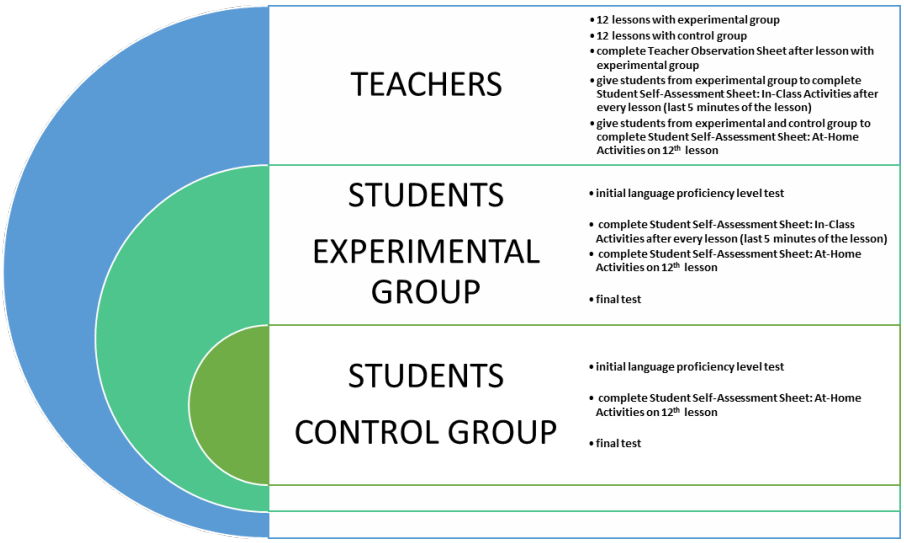

Figure 1. *Research protocol*

This research framework encompasses several different aspects of teaching. Initial and final testing provide data on language achievement and enable comparison of the control and experimental groups. The initial assessment establishes students' level of language knowledge before the instructional cycle, while the final test provides data on their achievement after its completion. Test data relate to language achievement but do not describe the course of the teaching process itself. According to the common research protocol, each teacher was expected to conduct 12 teaching hours with the control group and 12 teaching hours with the experimental group. The control group participated in teaching without the use of AI tools, whereas AI tools were incorporated into teaching activities in the experimental group. At the end of the instructional cycle (semester), both groups were expected to take

the same final test, covering the content taught during the course. The protocol also stipulated that the final tests administered by all project partners should have the same total number of points.

Monitoring of the teaching process was more detailed in the experimental group. After each teaching hour, the teacher was expected to complete the *Teacher Observation Sheet*, minutely detailed upon in sub-chapter 4.3.1 above, which recorded the implementation of classroom activities. At the same time, students in the experimental group completed the *Student Self-Assessment Sheet: In-Class Activities*, described in its characteristics in sub-chapter 4.3.2 above, during the final five minutes of each class. After completing the 12 teaching hours, students in both groups completed the *Student Self-Assessment Sheet: At-Home Activities*, referred to, regarding their architecture, in sub-chapter 4.3.3 above. Data were thus collected continuously throughout the instructional cycle and after its completion and include data on the teaching process from both the students' and teachers' perspectives.

The Teacher Observation Sheet was completed by the teacher after each class with the experimental group. Its purpose is to record the teacher's observations immediately after the activity, thereby providing data on the implementation of teaching from the teacher's perspective. The teacher recorded the course of the activity, the need for additional explanations, difficulties encountered during the work, student reactions, and any deviations from the planned implementation.

Because the teacher both conducted the activity and evaluated its implementation, these records represent a professional assessment of the teaching process rather than a fully independent measurement. Their research value is therefore enhanced when they are considered alongside other data

sources, particularly student self-assessments and language-test results. Comparing these data provides a more comprehensive view of the implementation of AI-supported teaching. Comparing student and teacher perspectives provides information on how students experience the use of AI tools and how teachers perceive their application during the teaching process.

Students completed the *Student Self-Assessment Sheet: In-Class Activities* after each class, providing insight into their immediate experience of using artificial intelligence tools. The student form was completed immediately after the teaching activities in order to capture students' specific experience of using AI tools. In this way, data were collected on their assessment of the usefulness of the tools, difficulties encountered during use, and how the tools were used in individual activities. Immediate data collection allows responses to refer to the activity just completed rather than only to a general impression of AI tool use over the entire instructional period. Student self-assessments serve a different research function from the measurement of language achievement. They are used to examine how students experience the use of artificial intelligence and their own participation in AI-supported activities, while language achievement is assessed using separate measurement instruments. A student's assessment that an AI tool was useful is a measure of student experience rather than an indicator of language progress. Assessments of autonomy, satisfaction, and the quality of generated responses are interpreted in the same way. Separating these types of data makes it possible to monitor student experience and language achievement independently. The data obtained illustrate why, when evaluating the use of artificial intelligence tools, perceived usefulness needs to be distinguished from how students report using AI tools in the learning process.

The final *Student Self-Assessment Sheet: At-Home Activities* form was completed by students in both the experimental and control groups after the 12 teaching hours. It collected data on the use of AI tools outside class. Such monitoring was relevant because AI tools were available to students outside organised teaching activities as well. This means that students in the control group could also use them independently, even though the use of artificial intelligence tools was not incorporated into their classes. Data on at-home activities made it possible to distinguish more clearly between the use of AI tools as part of planned teaching activities and their independent use outside class. This is particularly relevant when interpreting differences between the experimental and control groups. Membership in the control group meant that AI tools were not included in the organised teaching intervention, but it did not necessarily mean that students did not use AI tools in other learning situations. The final form also captured this aspect of their experience.

The observation sheets intended for teachers and students serve different research functions. The teacher form records teachers' observations of the implementation of activities in the experimental group, while the student form completed after each class provides insight into students' classroom experiences. The final questionnaire on at-home activities, completed by students in both groups, collects data on their activities outside the classroom. In this way, different, but interrelated information is obtained about the use of artificial intelligence tools during and outside class and about teachers' and students' experiences of their use. Test results provide data on language achievement, whereas student assessments provide insight into experiences associated with the use of AI tools. These are

different types of data, since a positive assessment of the experience of using AI tools does not necessarily imply better language achievement.

The common research protocol is applied across different teaching contexts while retaining the same basic elements of the research procedure. Because ESP teaching, learning and assessing at partner institutions concerns different professional fields (namely, education, mobility, criminology and justice, security, and tourism), teaching content and activities were adapted to the needs of individual student groups. The studies are therefore not based on completely identical teaching content, but on a common framework for its implementation and monitoring. The same basic research procedure (initial assessment of language knowledge, work with control and experimental groups, monitoring of the teaching process, and final assessment) is thus applied in different ESP contexts, with variation in teaching content, student-group characteristics, and the implementation of teaching activities. Therefore, although the specific content varied by field, all teams followed the common research protocol.

5.3 Overview of Research Studies Conducted as Part of the ELITE-AI Project

Section 3 of this chapter describes the research conducted by members of the project consortium. Project studies address different aspects of the use of AI tools in ESP teaching, learning and assessing. The published studies differ in their research aims and methodological approaches. Some compare teaching with AI tools and traditional teaching, some focus on student experiences and patterns of use, and others evaluate the linguistic,

pragmatic, and cultural appropriateness of content generated by artificial intelligence tools.

In the study by Gugić and Blažević (2026), both the control and experimental groups worked with the same teaching content. The experimental group used AI tools and had a greater degree of autonomy in carrying out activities, whereas the control group was guided to a greater extent by the teacher. In that study, the control group worked with pre-prepared teaching materials and teacher-led activities. The experimental group received initial instructions and expected outcomes but completed tasks more independently with the use of AI tools. The groups therefore differed not only in the use of AI but also in the degree of student autonomy and teacher guidance. Consequently, the results cannot be attributed exclusively to the use of AI tools but should be understood within the broader AI-supported learning approach. When the mode of work and the teacher's role differ alongside the use of AI tools, the effect of AI tool use cannot be clearly separated from the effects of other characteristics of the teaching process. These differences included the degree of teacher guidance and student autonomy, meaning that AI tool use was not the only difference between the experimental and control groups.

As mentioned in Section 2 of this chapter, the initial assessment serves to establish the students' initial level of language knowledge and to examine the comparability of the experimental and control groups. The data obtained provide a basis for interpreting the final results because initial differences in language knowledge may be associated with students' achievement after the instructional period. In the study by Gugić and Blažević (2026), initial testing was conducted at the beginning of the winter

semester of the 2025/2026 academic year, i.e., October 2025 and final testing at the end of the semester, in January 2026. The final test, designed according to the protocol of the ELITE-AI project described above, contained 40 items, 10 each for listening, reading, vocabulary, and grammar. The initial and final tests were not identical, but they measured the same language competence using different tasks. The data were analysed using descriptive statistics, an independent-samples t-test, and analysis of covariance (ANCOVA), with the initial score taken into account in the analysis of final achievement.

In the statistical analysis, the experimental group consisted of 22 female students and the control group of 18. The mean initial-test score was 22.73 points in the experimental group and 23.61 in the control group. On the final test, the experimental group achieved a mean score of 35.46, compared with 36.11 points in the control group. After initial knowledge was taken into account in the analysis, no statistically significant difference in final achievement was found between the groups. Under the conditions observed in this study, no advantage of the AI-supported approach over traditionally guided instruction was therefore identified in terms of final language achievement. This result does not imply that AI tools have no educational value; rather, it indicates that in this study their use was not associated with statistically significantly better final achievement. Characteristics of the research design are also relevant to the interpretation of the results, including the sample size and the fact that the experimental group differed from the control group in having greater student autonomy and less teacher guidance. The effect of AI tool use therefore cannot be separated in this study from the other characteristics of the AI-supported teaching approach.

Another aspect tackled concerns the use of AI tools in teaching. The experimental and control groups worked with the same teaching materials. The control group worked without AI tools, whereas the experimental group used them during teaching activities. The topics addressed were related to the students' future profession as early childhood educators and included multiple intelligences, learning styles, well-known educators, kindergarten work protocols, and planned grammar content. Teaching activities focused on listening and reading comprehension, vocabulary, and grammar. In the control group, listening instruction, for example, was divided into pre-listening, while-listening, and post-listening activities led by the teacher. The experimental group worked with the same vocabulary, but students used AI tools to identify the meanings of the target words. They then worked in small groups to use AI tools to create a text containing the target vocabulary, convert the text to speech, and formulate comprehension questions. A similar procedure was used for reading, grammar, and vocabulary activities.

Another important difference between the groups concerned the mode of work. According to the authors, the teacher prepared teaching materials in advance and led activities for the control group, whereas students in the experimental group worked more independently. They received initial instructions or topics and information about the expected outcome, but the way in which that outcome was achieved was left to them to a greater extent. The authors therefore explicitly state that the experimental condition differed from the control condition not only in the use of AI tools but also in the degree of student autonomy and teacher guidance. They consequently describe the findings as reflecting an AI-supported learning approach rather than an isolated effect of AI technology (Gugić & Blažević, 2026).

Nechifor (2026b) examined patterns of AI integration in ESP courses at four European universities participating in the ELITE-AI project. The study employed a descriptive quantitative design combining structured student self-assessment with teacher observation. The instruments focused on several dimensions of AI-supported learning, including the perceived usefulness of AI tools, students' active participation, help-seeking behaviour, independence in decisions concerning AI use, critical evaluation of AI-generated content, and the integration of AI into the broader learning process. By combining student and teacher perspectives, the study sought to examine not only students' attitudes towards AI tools but also patterns associated with their use during ESP instruction. The results indicated a generally positive perception of AI-supported activities. A total of 91.1% of student assessments indicated that students considered AI integration helpful. However, Nechifor (2026b) distinguishes this positive assessment from the ways in which students actually engaged with AI tools. In terms of autonomy, 66.3% reported using AI because their teacher had instructed them to do so, whereas 32.7% demonstrated greater independent initiative in its use. The findings therefore point to a distinction between perceived usefulness and independent engagement: students may evaluate AI-supported activities positively without necessarily initiating the use of AI tools independently. Critical evaluation of AI-generated content represented another important dimension of the study. Although 63.4% of students reported checking AI-generated content, 28.7% accepted generated responses without critically questioning their accuracy. These findings are particularly relevant in the context of ESP, where students work with specialised terminology and professional communication and where the reliability and appropriateness of generated information may

therefore be important. Positive attitudes towards AI tools do not in themselves demonstrate that students possess the critical AI literacy required to evaluate generated content systematically. The study consequently draws attention to the distinction between the ability to use an AI tool and the ability to assess its output critically.

A similar distinction emerged in relation to the integration of AI into the learning process. While 59.4% of students perceived AI as a useful classroom tool, 39.6%[4] reported using it primarily for individual tasks without clearly connecting these activities to the broader learning process. This finding suggests that the presence of AI in classroom activities does not necessarily imply its broader pedagogical integration. Nechifor (2026) therefore considers AI integration in relation to the extent to which individual AI-supported activities are connected with learning objectives and with the wider process of language learning, rather than simply in terms of whether an AI tool is used.

Taken together, these findings reveal a discrepancy between students' generally positive perceptions of AI and several aspects of their reported behaviour. Perceived usefulness coexisted with considerable teacher-directed use, incomplete critical evaluation of generated content, and the use of AI for isolated tasks. Nechifor (2026b) interprets these patterns as evidence that students' satisfaction with AI-supported activities and the

[4] The percentages reported by Nechifor (2026b) were calculated on the complete dataset, once data collection for the corresponding stage of the project had been concluded. They therefore differ from the figures cited in Chapter 7, Section 7.5.2 (54.5% directed use, 35.8% proactive use and 9.7% avoidance), which were drawn from the *Manual of Good Practices for Teachers and Students* (ELITE-AI Consortium, 2026b) at a point when data collection was still under completion. The difference between the two sets of figures should be read in light of the date at which each analysis was carried out.

pedagogical integration of AI represent different dimensions of the learning process. From a methodological perspective, this distinction is particularly relevant to the ELITE-AI research framework because student self-assessment data provide information about students' experiences and reported patterns of AI use, rather than direct evidence of improved language achievement.

Smajla and Podovšovnik (2026) likewise show that positive student attitudes towards AI do not necessarily mean that AI tools are always used appropriately. Their findings indicate that students use AI tools for different purposes, most often for completing assignments, improving written expression, and finding new information. Although students recognise the benefits of such tools, they also express concerns about possible over-reliance on artificial intelligence, reduced independent and critical thinking, and the unreliability of generated information. The results show that a positive assessment of the usefulness of AI tools does not in itself provide information about how they are used or about achieved language outcomes. Student self-assessments within the project research framework therefore represent a distinct type of data from teacher observations and the results of initial and final assessment.

The next study conducted within the framework of the ELITE-AI project is the one by Smajla and Kompara Lukančič (2025). Its findings show that the use of AI is not limited to organised teaching activities. In a study conducted among students in Slovenia, students reported using AI tools for various activities related to learning languages for specific purposes, particularly for preparing assignments, improving written expression, correcting grammatical errors, and acquiring new knowledge and skills.

Students also reported potential problems, especially over-reliance on AI tools, weakening of independent thinking, and the unreliability of the information obtained. The findings indicate that students' experience of using artificial intelligence tools may also be shaped by their use outside the classroom. The final at-home activities form therefore made it possible to take independent AI tool use into account when interpreting project data, rather than considering only their use during teaching activities.

A separate strand of research concerns the automatic processing of specialised language material. Kompara Lukančič and Smajla (2025) analysed the automatic recognition of English abbreviations and their full forms in specialised texts in the fields of criminology, justice, and security using the *Krajšavar* algorithm. This work was followed by a study by Kompara Lukančič (2026a) focusing on the automatic recognition of abbreviations and their full forms in Italian scientific texts in the field of tourism. In that study, the texts were analysed using the *Krajšavar* algorithm and the AI tool *Copilot*, and the results of their application were compared with the aim of improving the recognition of Italian abbreviations and their full forms. These studies extend the project's examination of AI tools to the processing of specific lexical and terminological features of specialised language.

Project studies also addressed the quality and appropriateness of content generated by AI tools for ESP learning. Nechifor, Turculeț, and Vasilescu (2025) examined AI-generated learning scenarios designed for students preparing for international mobility. The scenarios were produced through students' interactions with five GenAI platforms – *ChatGPT, Claude, Copilot, DeepSeek,* and *Gemini* – and covered four types of communicative

situations: multicultural learning environments, social pragmatics, administrative situations, and linguistic pragmatics. Students were required to incorporate specialised vocabulary and appropriate grammatical structures into the scenarios, thus placing the evaluation of AI-generated content within communicative situations relevant to academic mobility and ESP learning. The study applied a multidimensional evaluation framework grounded in ESP, intercultural competence, and pragmatics. Five main dimensions were assessed: pragmatic authenticity, linguistic appropriateness, cultural sensitivity, discourse coherence, and pedagogical authenticity. This framework extended the evaluation beyond grammatical correctness by examining whether AI-generated scenarios reflected appropriate speech acts and pragmatic choices, suitable linguistic register, culturally sensitive representations, coherent discourse, and meaningful pedagogical value. The analysis therefore treated the quality of AI-generated ESP materials as a multidimensional construct rather than reducing it to linguistic accuracy alone.

The methodological design combined computational and human evaluation. Approximately 200 AI-generated ESP scenarios were processed through an automated Python-based evaluation system using the locally installed *Llama 3.1* model within the *Ollama* framework. The computational analysis included the identification of linguistic and pragmatic features such as hedging and mitigating devices, pragmatic markers, specialised vocabulary, and genre-related features. *Llama 3.1* subsequently evaluated the scenarios according to the five-dimensional framework and a more detailed set of 13 sub-criteria. The automated analysis was complemented by evaluations conducted by four human specialists in applied linguistics, who assessed the outputs using the same

major evaluation dimensions. This made it possible to examine the extent to which computational assessment corresponded with expert human judgement.

The findings revealed differences both across evaluation dimensions and among the five GenAI platforms. Cultural sensitivity received the highest overall mean score (4.55), followed by pedagogical value (4.16) and pragmatic authenticity (4.03). Lower scores were obtained for discourse coherence (3.92) and linguistic appropriateness (3.88), indicating that register management, conversational naturalness, and aspects of discourse organisation remained more problematic. The results therefore show that comparatively strong performance in cultural sensitivity or pedagogical usefulness did not necessarily imply equally strong performance in all linguistic and discourse-related dimensions.

Platform-specific differences were also identified. *Gemini* obtained the strongest results across most of the evaluated dimensions, while the other platforms displayed different profiles of strengths and limitations. The authors report that *ChatGPT* showed reliable cultural sensitivity and pedagogical structure, *Claude* performed particularly well in discourse organisation but required closer attention to linguistic consistency, while *Copilot* and *DeepSeek* showed more moderate or basic performance patterns. At the same time, difficulties related to linguistic register and conversational naturalness occurred across platforms, indicating that these aspects remained challenging even when the overall quality of generated scenarios was relatively high. An additional component of the study compared automated assessment with expert human evaluation. The relationship between *Llama 3.1* evaluations and human specialist ratings

was examined using Spearman's rank correlation. The analysis produced a moderate positive correlation ($\rho = .40$), but this relationship was not statistically significant ($p = .51$). The authors interpret this result cautiously because the validation involved only five platforms and four human evaluators. The finding therefore indicates partial correspondence between computational and expert assessments rather than demonstrating that automated evaluation can substitute for human judgement. Taken together, the results demonstrate that the evaluation of AI-generated ESP materials involves several distinct dimensions of quality. Content may be culturally sensitive and pedagogically useful while still presenting weaknesses in linguistic appropriateness, discourse coherence, or register management. Nechifor, Turculeț, and Vasilescu (2025) consequently emphasise analytical evaluation and critical human supervision of AI-generated materials, particularly when such content is intended for specialised and intercultural communication. Their comparison of computational and human evaluations further supports the use of human expertise alongside automated assessment rather than reliance on computational evaluation alone.

Costantini, Gallitelli, and Li Destri (2026) provide an overview of the opportunities and challenges associated with the use of AI tools in education, with particular attention to their application in ESP teaching, learning and assessing. Their analysis is based on a review of 100 publications identified through targeted Google Scholar searches conducted between November 2024 and January 2025. The authors explicitly note that the review did not follow a fully systematic protocol; studies were selected according to keyword relevance, citation visibility, and thematic alignment. The review should therefore be understood as a

synthesis of dominant trends and recurring issues in recent literature rather than an exhaustive or statistically representative account of the field. Among the main opportunities identified in the literature are personalised and adaptive learning, increased learner autonomy and engagement, greater efficiency of particular educational processes, and support for skill development. At the same time, the review identifies recurring concerns related to academic integrity, over-reliance on AI tools, the accuracy and reliability of generated content, algorithmic bias, data privacy, reduced human interaction, and unequal access to technological infrastructure. In the specific context of ESP, the authors observe that research remains largely exploratory and relies strongly on perception-based studies. They also identify the limited use of mixed-methods designs as a methodological constraint, particularly because it restricts conclusions about sustained and long-term learning benefits. The authors consequently present AI as a tool that can support teaching and learning rather than replace the teacher. They emphasise the continuing role of the teacher as a critical facilitator and identify digital and AI literacy as important elements of responsible AI integration. This includes understanding the functioning and limitations of AI tools, recognising their ethical implications, and critically evaluating AI-generated content (Costantini, Gallitelli, & Li Destri, 2026).

5.4 Conclusion

Project studies have addressed different aspects of the use of AI tools in the field of ESP. Gugić and Blažević (2026) compare AI-supported and traditional teaching approaches and analyse the language achievement of experimental and control groups. Smajla and colleagues investigate student

attitudes and patterns of AI tool use in learning languages for specific purposes, while Nechifor (2026b) examines students' experiences with AI-supported ESP learning, focusing on perceived usefulness, autonomy in AI tool use, critical evaluation of AI-generated content, and its integration into the learning process. Nechifor, Turculeț, and Vasilescu (2025) evaluate the linguistic, pragmatic, and cultural appropriateness of content generated by GenAI platforms. Costantini, Gallitelli, and Li Destri (2026) place these issues within the broader context of recent research on the possibilities and challenges of AI tool use in education and ESP.

An additional research perspective is provided by studies focusing on the automatic processing of specialised language material. Kompara Lukančič and Smajla (2025) investigate the automatic recognition of English abbreviations and their full forms in texts from the fields of criminology, justice, and security using the *Krajšavar* algorithm. Kompara Lukančič (2026a) extends this research to Italian scientific texts in tourism and compares the results of the *Krajšavar* algorithm and *Copilot* in recognising abbreviations and their full forms. Alongside classroom use and student experiences, project studies thus also address the potential use of computational and AI tools in analysing specific features of specialised language.

The studies analysed therefore do not provide a single general answer regarding the effectiveness of AI tools in ESP teaching and learning; rather, they provide evidence on different aspects of their use: the organisation of AI-supported teaching activities, language achievement, student attitudes and patterns of use, the quality of AI-generated content, and the automatic processing of specialised language material. This range of research

questions demonstrates that the use of artificial intelligence tools in English for Specific Purposes can be examined at multiple levels, with the findings of each study interpreted in relation to its aims, methodological procedures, and type of data collected.

The findings of the project studies indicate the need to distinguish students' attitudes towards AI tools from indicators of language achievement. Positive assessments of the usefulness of AI tools provide information about students' experiences and attitudes but do not in themselves confirm better language-learning outcomes. Conversely, the absence of a statistically significant difference in final language achievement does not indicate how students experience AI tools, the purposes for which they use them, or how frequently they use them. The studies analysed therefore illuminate different dimensions of the use of AI tools in ESP teaching, learning and assessing, each arising from different research questions and methodological procedures.

CHAPTER SIX

Analysis of AI-Integrated Language Teaching Practices: Developed Concepts and Good Practice Models

6.1 An Overview

The integration of Artificial Intelligence (AI) into language education has become a central focus for researchers (some recent literature overviews: i.e. Mohebbi, 2025; Son et al., 2025; Zhu & Wang, 2025). Within the framework of WP3 of the ELITE-AI project, partner institutions implemented a wide range of classroom-based activities designed to explore how AI tools can support language learning and improve students' performance, while at the same time supporting critical engagement with digital technologies. The experimental lessons collected across institutions represented diverse disciplinary contexts such as tourism, education, criminal justice, and academic mobility and were implemented across a range of classroom settings involving university-level learners with varying levels of proficiency, from first-year undergraduate to postgraduate students. The lessons varied in duration, typically lasting between 45 and 100 minutes, depending on institutional scheduling. A variety of AI tools were employed throughout the activities, including general-purpose chatbots (e.g., *ChatGPT, Copilot, Gemini,* and *Perplexity*), as well as more specialised platforms, i.e. for grammar checking. These tools were integrated into the teaching process to support the development of a wide range of language skills, including reading, writing, speaking, listening, vocabulary, and grammar. Despite this diversity, the activities share a

common objective: to integrate AI tools into language teaching in a structured and pedagogically meaningful way.

This section presents a synthesis of the results of these classroom implementations, focusing specifically on two key aspects: the conceptual roles of AI emerging from practical use, and the examples of good practice reflected in the design and execution of learning activities. Rather than offering a descriptive account of individual lessons, the analysis aims to identify recurring patterns and to systematise them into transferable models. The goal is to move from isolated teaching practices to a structured understanding of how AI can be effectively embedded in language curricula. While a significant body of existing research has focused on single classroom experiences or course-specific implementations (i.e., Boeru, 2024; Dou, 2024; Silitonga, 2024), even in broader studies there remains a relative lack of approaches that combine multiple field experiences with a systematic analysis based on their abstraction. The chapter is therefore organised into three main components. First, a conceptual framework outlines the primary pedagogical functions attributed to AI across the various contexts. Second, a typology of AI-integrated activities presents concrete examples of good practice that can be replicated or adapted by educators. Third and final, a cross-context analysis highlights key patterns in implementation, focusing on factors that influence the effectiveness of AI-supported learning.

6.2 Roles of AI in Language Teaching

The analysis of the experimental lessons conducted across partner institutions makes it possible to identify a set of recurring conceptual roles

attributed to AI tools in language teaching. Although the activities varied in terms of disciplinary context and proficiency levels, the ways in which AI was integrated into classroom practices reveal a shared pedagogical direction. This section outlines the main conceptual roles of AI that can be derived from the collected teaching practices. These roles do not represent isolated uses, but rather interconnected functions that often coexist within the same activity.

6.2.1 AI as a Drafting and Revision Tool

One of the most prominent roles of AI across the analysed lessons is its function as a drafting and revision tool within process-oriented writing activities. In this context, AI does not replace the learner's productive effort but supports the development of written language through iterative refinement. Students frequently engaged with AI tools to generate initial drafts of texts such as emails, reports, motivation letters, or descriptive passages. However, these outputs were rarely treated as final products. Instead, they served as a starting point for further analysis and revision. Students were often required to contrast AI-generated content with their own writing or with model texts, identifying differences in structure, vocabulary, tone, and accuracy. This process-oriented use of AI reflects a shift away from product-focused writing towards a more reflective approach. The pedagogical emphasis is placed on recognising the strengths and weaknesses of AI output and making informed linguistic choices. Importantly, the revision phase emerges as the most significant learning moment, as it is precisely at this stage that learners actively engage with language, developing metalinguistic awareness and critical judgement. AI thus becomes a catalyst for reflection rather than an endpoint of production.

6.2.2 AI as a Task Generator and Learning Resource Designer

A second key conceptual role is that of AI as a generator of learning materials. Unlike traditional instructional models, where exercises are designed exclusively by teachers or textbook authors, AI enables both educators and students to create customised practice tasks. In many of the lessons, students used AI tools to generate grammar exercises, vocabulary activities or comprehension questions. This shift transforms the learner's role from passive recipient to active creator of learning content. In order to obtain useful outputs, students must formulate clear and specific prompts, which in turn requires a deep understanding of the target language. From a pedagogical perspective, this use of AI introduces several advantages, particularly in increasing learner autonomy and allowing for the adaptability of materials to individual needs. These benefits have been identified as central in studies such as Ningsih et al. (2023) and Du and Alm (2024). Furthermore, the methods observed in the lessons create opportunities for peer-to-peer learning, as students can share and evaluate the tasks they generate. Moreover, the act of designing exercises reinforces language awareness. Students must consider not only correct usage and typical errors, but also the appropriateness of their linguistic choices for the target context and the difficulty level. In this sense, AI supports a form of 'learning by teaching', where generating tasks becomes an integral part of the learning process.

6.2.3 AI as a Mediation Tool between Learners and Authentic Input

Another significant role of AI identified across the dataset is its function as a mediator between learners and authentic materials. Rather than simplifying or replacing complex texts, AI provides on-demand support

that enables students to engage more effectively with real-world language. Learners used AI tools to obtain explanations of unfamiliar vocabulary, generate paraphrases or simplified versions of complex sentences, request examples in different contexts, and clarify meanings in discipline-specific texts. This type of use also reflects patterns reported in student questionnaire-based studies, such as Synekop et al. (2024), which highlight that learners commonly rely on chatbots for vocabulary explanation and clarification. Introducing these practices in the classroom, under teacher supervision, makes it possible to integrate already existing student behaviours into structured learning environments, while at the same time promoting awareness of good practices in AI use.

The use of AI tools as mediators allows students to maintain exposure to authentic input while reducing the cognitive load associated with comprehension. Instead of interrupting the learning process to consult multiple external resources (e.g., dictionaries, grammar references), learners can interact dynamically with the text through AI. Crucially, this mediation function promotes learner autonomy. Students decide what they need to understand and use AI as a tool for targeted inquiry. As a result, comprehension becomes a more active and personalised process. At the same time, the use of AI in this capacity highlights the importance of guiding students in evaluating the reliability of AI explanations. Without appropriate scaffolding, there is a risk of over-reliance on AI interpretations, particularly in specialised domains.

6.2.4 AI as a Trigger for Critical Evaluation and Metacognitive Development

Beyond its supportive functions, AI also acts as a powerful trigger for critical evaluation and metacognitive development. In many of the activities, the pedagogical value did not lie in the correctness of AI output, but in its limitations. Students were frequently exposed to AI-generated texts that contained inaccuracies or inappropriate terminology. Rather than correcting these issues directly, teachers designed tasks that required learners to identify and analyse such problems. This approach facilitates a critical stance toward AI and encourages deeper engagement with both language and content.

Indeed, through these activities, learners developed the ability to question the reliability of AI-generated information and to verify outputs against authoritative sources. They also learned to detect inconsistencies or ambiguities and to justify both linguistic and factual corrections. This shift from passive consumption to active evaluation represents one of the most significant pedagogical impacts of AI integration. It aligns with broader educational goals related to digital literacy that involves critical thinking and responsible technology use.

Moreover, critical engagement with AI encourages metacognitive reflection. Students become more aware of their own learning processes, recognising when to trust AI output. This awareness is particularly important in academic and professional contexts, where accurate communication is essential.

6.2.5 AI as Support for Genre-based and Professional Language Learning

Across the lessons, AI was consistently used to support the learning of specific genres and professional communication practices. Rather than focusing exclusively on general language proficiency, many activities were centred on producing texts relevant to real-world contexts. These included:

i. formal emails and letters;
ii. reports and briefings;
iii. informational leaflets;
iv. legal or institutional documents;
v. FAQ-style explanatory texts.

AI tools facilitated access to these genres by providing examples, templates, or lexical suggestions. However, students were required to adapt these outputs to match specific communicative purposes and audience expectations.

This use of AI reflects a shift towards more contextualised and purpose-driven language learning. Instead of practicing isolated linguistic forms, learners engage with language as it is used in professional and institutional settings. An important aspect of these activities is the attention to register and audience. Students learn to adjust tone, vocabulary, and level of detail depending on the target reader, moving beyond correctness towards appropriateness. AI supports this process by offering alternative formulations, but the responsibility for selecting these options remains with the learner.

6.2.6 Dynamic and Context-dependent Roles of AI

Finally, it is important to emphasise that the roles of AI described above are not fixed or mutually exclusive. In most activities, AI operates simultaneously in multiple capacities, and its function evolves depending on the task design. For example, in a single lesson, AI may simultaneously act as a drafting tool by generating initial texts, a mediation tool by providing vocabulary explanations, a task generator by producing additional exercises, and a feedback tool by offering alternative formulations for revision. The effectiveness of AI integration therefore depends not on the tool itself, but on how it is embedded within pedagogical sequences. Structured tasks and clear objectives, along with guided reflection, are essential for ensuring that AI contributes meaningfully to learning outcomes.

A key finding across contexts is that the most successful implementations were those in which AI use was purposeful rather than incidental, integrated into task design rather than added as an optional resource, and consistently accompanied by critical evaluation and discussion. This highlights the central role of the teacher as a designer of learning experiences. While AI expands the range of available resources, it does not replace the need for pedagogical expertise. On the contrary, it increases the importance of careful task planning and instructional guidance.

6.3 Typology of AI-Integrated Activities

The experimental teaching activities conducted across partner institutions reveal a set of recurring pedagogical patterns in the integration of AI tools

into language instruction. While the thematic content and institutional contexts varied (ranging from criminal justice and tourism to education and academic mobility), the structure of classroom activities showed a significant degree of convergence. These shared practices can be synthesised into a typology of AI-integrated activities, representing transferable models of good practice suitable for implementation across diverse educational settings. The following subsections present these activity types, describing their pedagogical structure, learning objectives, and representative examples drawn from the experimental lessons.

6.3.1 AI-Assisted Writing with Iterative Revision

One of the most consistent and pedagogically robust activity types identified across the dataset is AI-assisted writing structured around iterative revision cycles. In this model, AI is not used as a final content generator but as a tool supporting multiple stages of the writing process. The activity typically unfolds in four stages:

i. initial drafting, often supported by AI-generated prompts or text;
ii. comparison between AI output and student-produced content;
iii. revision based on identified discrepancies;
iv. feedback through peer review or teacher intervention.

This approach was widely observed across disciplines. For example, in tourism courses, students used AI tools such as *Copilot* and *ChatGPT* to draft professional emails and cover letters, subsequently revising them in light of peer feedback and formal writing guidelines. Similarly, in academic mobility classes, students generated draft motivation letters and then compared these with their own writing, identifying differences in tone,

structure, and lexical choice before engaging in a revision phase. In criminal justice courses, writing tasks such as police reports or institutional briefings followed a similar pattern: AI was used to generate model sentences or structures, but learners were required to critically assess these outputs and ensure factual and terminological accuracy.

This activity type supports several key learning outcomes. It reinforces process-oriented writing and encourages metalinguistic awareness: this positions AI as a resource for reflection rather than as a substitute for learner effort. The revision stage is indeed central: the pedagogical value lies not in the initial AI output but in the learner's engagement with it. Moreover, given that learners are likely to rely on AI tools in their future professional contexts to produce similar types of texts, it is essential that they develop critical skills at this stage in order to evaluate such outputs effectively and ensure a responsible use of AI.

6.3.2 AI-Generated Practice and Student-Designed Tasks

A second major activity type involves the use of AI tools as generators of practice material, often under direct student control. In contrast to traditional teacher-led exercise design, learners actively engage in task creation by prompting AI systems to generate grammar exercises, vocabulary tasks, or comprehension questions. This activity typically follows a learner-driven cycle in which students first identify target language features, such as a grammatical structure or a set of vocabulary items, then prompt the AI tool to generate exercises or examples, and finally complete the generated tasks individually or exchange them among peers.

This model was particularly evident in grammar-focused sessions in the mobility curriculum, where students used AI tools to create gap-fill exercises, transformation tasks, or short quizzes based on previously introduced vocabulary and grammar points. Similarly, in tourism classes, students generated customer enquiries using AI tools and then used these as prompts for writing formal responses. In education lessons, students working on present tenses asked AI tools to generate test questions and example sentences without explicit rules, from which they independently inferred grammatical patterns before confirming them, turning task generation into a form of guided discovery learning.

The pedagogical value of this approach lies in its inversion of traditional roles. Students move from passive recipients of material to active task designers, which requires a deeper understanding of language structures and usage. Moreover, prompting the AI becomes a linguistic activity in itself, requiring learners to formulate precise instructions in order to obtain useful outputs. This model also supports differentiation, as students can generate tasks at varying levels of difficulty and complexity, tailored to their individual needs.

6.3.3 AI-Supported Role-Play and Simulation Tasks

Another highly effective activity type identified across the dataset is the use of AI in role-play and simulation-based tasks. These activities aim to replicate professional or real-world communicative scenarios, with AI providing support in either preparation/execution or feedback phases. Typical scenarios include job interviews or academic mobility applications, police interrogations or legal procedures, and customer communication in tourism settings. For example, students in criminal justice courses designed

simulated interrogation scenarios, using AI tools to generate procedural elements such as personal profiles and dialogue structures. In mobility classes, role-plays simulated interview situations, with AI supporting both the preparation of questions and the evaluation of student responses.

In these cases, AI often played a dual role, functioning both as a preparatory tool by generating scenarios, vocabulary, or dialogue structures, and as a feedback provider by offering comments on register and accuracy. The integration of AI into simulation tasks significantly extends the reflective dimension of speaking activities. For instance, students could transcribe or summarise their spoken interactions and submit them to AI for feedback, thereby transforming ephemeral oral performance into analysable textual data. This activity type promotes communicative competence, focusing on functional language use and professional discourse awareness. It also enables scalable feedback in large classes, where individual teacher feedback would otherwise be limited.

6.3.4 AI-Mediated Reading and Vocabulary Development

A further key activity type involves the integration of AI tools into reading comprehension and vocabulary acquisition processes. In this model, AI acts as a mediation layer between learners and authentic texts. Activities typically involve reading authentic or semi-authentic texts, using AI to clarify unfamiliar vocabulary, requesting paraphrases or simplified explanations, and generating example sentences or collocations.

In mobility courses, students engaged with texts about Erasmus+ programmes and used AI tools to obtain contextualised definitions and paraphrases of complex expressions. Similarly, in criminal justice lessons,

learners used AI to summarise institutional descriptions and compare them with official documents. Education students also used AI tools to explore theoretical concepts such as learning styles by independently retrieving definitions and examples, rather than relying on pre-selected materials. This approach allows students to maintain engagement with authentic materials without being overwhelmed by linguistic complexity. Rather than replacing texts with simplified versions, AI provides adaptive support that can be adjusted to individual needs.

An important aspect of this activity type is the shift from teacher-mediated explanation to learner-driven inquiry. Students decide which elements require clarification and actively seek assistance from AI tools, promoting autonomy and strategic learning. Moreover, interacting with AI may be perceived as a safer option, as it is free from the evaluative judgement of teachers or peers, a factor that has been highlighted in studies such as Erito (2023) and Du and Alm (2024)

6.3.5 AI vs Authoritative Sources Comparison Tasks

One of the most pedagogically distinctive activity types involves the explicit comparison between AI-generated content and authoritative sources. In these tasks, students critically evaluate AI outputs by verifying them against official sources or course materials. The typical structure of these activities involves generating content using AI, such as summaries, definitions, or descriptions, and then retrieving corresponding information from verified sources. This is followed by identifying discrepancies or inaccuracies and engaging in discussion about them.

This model was prominently used in lessons focusing on institutional knowledge in criminal justice and security studies. For example, students generated summaries of national agencies or international organisations using AI tools and then compared these with official descriptions, paying particular attention to terminology and accuracy. These comparison tasks revealed frequent issues such as outdated information and terminological imprecision; particular attention was also paid to oversimplified explanations. In mobility classes, this approach was carried on in activities involving institutional documents such as Learning Agreements, where students used AI-generated drafts as a starting point but systematically compared them with official templates and guidelines, highlighting discrepancies in terminology or formal requirements.

Rather than treating AI limitations as obstacles, the activity design leveraged them as learning opportunities. Students developed critical literacy skills by identifying and correcting errors, thereby deepening their understanding of the subject matter. This activity type is particularly valuable in domains where factual accuracy and precise terminology are essential, as it encourages learners to adopt an analytical stance toward AI-generated content.

6.3.6 Genre Transformation and Audience Adaptation Tasks

The final activity type identified across the dataset involves the transformation of content across different genres and the adaptation of language to specific audiences. Suggestions by AI tools were used by students to modify texts according to communicative purpose and target readership. Typical tasks include transforming informational texts into formats such as reports, leaflets, or FAQs, adapting technical descriptions

for non-specialist audiences, and rephrasing content to accommodate different levels of formality and communicative contexts. For instance, students comparing EUROPOL and INTERPOL were asked to create FAQ-style texts addressing common misconceptions, requiring them to simplify complex institutional information for a general audience. Similarly, when studying NATO-related materials, students produced audience-specific explanations tailored to different professional groups, such as police officers or civilian personnel, adjusting both vocabulary and level of detail. In tourism courses, students transformed descriptive content into promotional leaflets, using AI-generated suggestions as a basis for revision and refinement. In mobility activities, students adapted the same content across different communicative purposes, for instance transforming formal institutional information into student-friendly explanations or guidance texts for incoming participants, adjusting both register and level of detail. This activity type highlights the importance of genre awareness and pragmatic language use. AI supports the generation of initial structures and vocabulary, but learners are responsible for ensuring appropriateness, coherence, and contextual relevance.

6.4 Cross-Context Patterns in Activity Implementation

The analysis of AI-integrated teaching activities across partner institutions reveals a set of recurring patterns that transcend disciplinary, linguistic, and institutional differences. While the specific tasks varied in format and content, their implementation highlights shared dynamics in learner behaviour and task effectiveness. These cross-context patterns provide valuable insights into how AI can be most effectively integrated into language teaching and

learning: they mainly concern the relationship between task type and learner engagement, the development of critical AI use, the role of task design and prompting, and the evolving role of the teacher.

6.4.1 Variation in Effectiveness across Task Types

One of the most evident patterns emerging from the analysis concerns the variation in effectiveness of AI integration depending on the type of task. While AI was used across a wide spectrum of activities, its pedagogical impact was not uniform. In particular, activities involving communicative and context-rich tasks consistently showed higher levels of student engagement and more meaningful interaction with AI tools. These included activities such as role-play simulations, professional writing tasks and content creation for tourism or institutional contexts.

In such activities, students demonstrated greater autonomy and creativity, along with willingness to experiment with language. AI was used not only to generate content but also to improve it. Learners actively engaged in iterative processes, making decisions about what to accept, modify, or reject. By contrast, in more controlled tasks, such as grammar exercises, the interaction with AI tended to be more passive. Students were more likely to accept AI-generated outputs without critical evaluation or to use the tool merely to obtain correct answers. In some cases, this led to reduced use of the target language during the activity, as learners relied heavily on AI to complete tasks efficiently rather than reflectively.

This contrast suggests that AI integration is most pedagogically effective when it is embedded in tasks that require higher-order thinking and contextualised language use. Activities that position learners as active agents rather than answer-seekers create more opportunities for meaningful learning.

6.4.2 Development of Critical AI Literacy

A second major pattern concerns the progressive development of critical awareness in the use of AI tools. Across multiple contexts, students initially approached AI outputs with a relatively high level of trust, often assuming reliability. However, this tendency shifted significantly in activities that explicitly required evaluation of the outputs. Tasks that involved identifying inaccuracies or correcting terminological/factual errors proved particularly effective in developing critical AI literacy. In these contexts, students began to recognise the limitations of AI, including issues such as oversimplification, which often implies a lack of precision.

Importantly, this shift was not automatic but depended on task design. When activities explicitly incorporated evaluation phases, students developed a more sceptical approach. They began to question AI output and seek confirmation from reliable sources, in order to justify their corrections. This progression from trust to verification represents a key learning outcome. It reflects the development of digital literacy skills that extend beyond language learning and are essential in academic and professional environments. Moreover, critical engagement with AI contributed to deeper content learning. By identifying and correcting errors, students engaged more thoroughly with the subject matter, especially in specialised domains such as criminal justice or institutional communication.

6.4.3. The Role of Prompt Design and Task Constraints

Another consistent pattern across the analysed lessons is the importance of prompt design and task structuring in determining the quality of AI-supported learning. Effective activities were characterised by clearly

defined objectives and specific instructions regarding AI use; usually they involved structured sequences such as draft → compare → revise.

In these cases, AI functioned as an integrated component of the task rather than as an optional tool. Students were guided not only in what to do, but also in how to interact with AI systems. By contrast, less structured activities, where AI use was left largely unguided, often resulted in superficial engagement. Students tended to rely on generic prompts, leading to equally generic outputs. Without clear expectations, the interaction with AI remained limited in depth and pedagogical value. An important aspect of task design was the level of constraint applied to prompting. Activities that required students to formulate precise, context-specific prompts led to more meaningful outputs and deeper linguistic engagement. In this sense, prompting itself emerged as a key skill, involving, other than clarity of instruction, the awareness of context and the understanding of desired outcomes.

However, the data also suggest that learners benefit from scaffolding in this area. Providing examples of effective prompts or partially structured input can support students in developing this competence, particularly in the early stages of AI use. Overall, these findings highlight that the effectiveness of AI integration depends less on the technology itself and more on the pedagogical framework within which it is embedded.

6.4.4. Evolution of the Teacher's Role

The integration of AI tools into classroom activities also led to observable shifts in the role of the teacher. Rather than reducing the importance of the instructor, AI use appears to redefine and, in some respects, expand the

scope of teaching responsibilities. Traditionally, language teachers have played a central role in providing linguistic input, by correcting errors and offering explanations. While these functions remain relevant, the integration of AI introduces new responsibilities that are more aligned with guidance.

In the analysed lessons, teachers increasingly acted as:

i. designers of learning experiences, creating structured tasks that integrate AI meaningfully;
ii. facilitators, guiding students in the use of AI tools and supporting collaborative work;
iii. mediators of knowledge, helping students interpret and evaluate AI-generated content;
iv. critical guides, encouraging reflection on the reliability and limitations of AI.

One particularly important aspect of this evolving role is the teacher's function as an epistemic guide. In activities involving specialised content, teachers supported students in distinguishing between accurate and inaccurate information, especially when AI outputs were misleading or incomplete.

This shift also involves a redistribution of responsibility. Students take a more active role in generating content and identifying errors, while the teacher supports higher-level processes such as evaluation and reflection. Far from being marginalised, the teacher becomes central to ensuring that AI use leads to meaningful learning rather than superficial task completion.

6.4.5. Conditions for Effective AI Integration

Across all contexts, certain conditions consistently emerged as enabling factors for successful AI integration. AI should be purposefully integrated into tasks with clear pedagogical objectives, rather than treated as an add-on or novelty. Equally important is the design of structured activities that include defined stages, particularly those involving revision or evaluation, while requiring active and critical engagement with AI outputs, grounding tasks in authentic contexts, and supporting learners through guided use of tools with appropriate scaffolding and modelling. When these conditions were present, AI contributed positively to learner engagement and autonomy and appeared to aid skill development. Conversely, their absence often resulted in limited or superficial use of AI.

6.5 Implications for Future Practice

The analysis of AI-integrated teaching practices across partner institutions provides a number of important implications for the future development of language education and teacher training. These implications concern not only the use of AI tools themselves, but also broader pedagogical considerations related to task design, learner engagement, and the development of digital and linguistic competences.

6.5.1. From Tool Adoption to Pedagogical Integration

One of the most significant findings is that the effectiveness of AI in language teaching depends less on the choice of tool and more on the way it is integrated into pedagogical design. Activities in which AI was introduced as an optional element tended to produce limited learning

outcomes. By contrast, when AI was embedded within clearly structured tasks with defined objectives, its contribution to learning became substantially more meaningful. This suggests that future practice should move beyond a focus on 'using AI tools' towards a more comprehensive approach centred on pedagogical integration. Educators need to design activities in which AI plays a functional role within the learning process, rather than serving as a generic support resource.

6.5.2. Importance of Task Design

The data strongly indicate that task design is a critical factor in determining the success of AI-supported activities. Effective tasks share several characteristics, such as clear sequencing and specific instructions regarding AI use. Future teaching practices should therefore prioritise the development of structured activity models that guide students through meaningful interactions with AI. Open-ended or unstructured use of AI tools, while potentially engaging, often lacks the depth required for sustained learning. In addition, activities should incorporate stages that require students to engage actively with AI outputs, rather than simply generating responses.

6.5.3. Developing Responsible AI Use

A central implication concerns the need to foster critical AI literacy as an integral component of language education. The analysis demonstrates that students do not automatically question the reliability of AI-generated content. Without explicit guidance, there is a tendency to accept outputs at face value. Future practice should therefore include tasks requiring verification of AI outputs against reliable sources and activities focused on identifying factual/linguistic inconsistencies, followed by guided

discussions on the limitations and appropriate use of AI. In this way, AI can become a powerful tool for developing not only language proficiency but also critical thinking and digital competence. These skills are increasingly essential in academic and professional contexts, where the ability to evaluate information is as important as the ability to produce it.

6.5.4. Supporting Genre Awareness and Professional Communication

The use of AI in the analysed lessons highlights its potential for supporting genre-based learning and professional communication. Activities involving emails, reports, institutional texts, or promotional materials allowed students to engage with language as it is used in real-world contexts. Such an approach aligns language learning more closely with employability skills and practical communication needs.

6.5.5. Teacher Training

The integration of AI also has important implications for teacher training. As observed in the analysed lessons, the role of the teacher evolves significantly in AI-supported environments. Teachers are required to design complex, multi-stage tasks and guide students in the effective use of AI tools. Moreover, they should both support critical evaluation of outputs and manage the balance between autonomy and guidance. This expanded role requires targeted training: teachers need opportunities to develop digital competence in the use of AI tools and experiment with activity design, while sharing practices and reflecting on outcomes. Providing educators with structured resources, such as the activity models outlined in this chapter, is therefore essential for sustainable implementation.

6.5.6. Enhancing Learner Autonomy

Finally, the integration of AI offers significant opportunities for enhancing learner autonomy. By allowing students to seek explanations and revise their work independently, and even generate materials, AI supports more self-directed learning processes. However, autonomy should not be understood as the absence of guidance. Rather, it should be supported through clear frameworks for task completion and scaffolding in prompt design and tool use, alongside reflective activities that encourage awareness of learning strategies. In this sense, AI can facilitate a shift towards more learner-centred approaches, provided that it is embedded within a supportive pedagogical structure.

6.6. Conclusion

The analysis of AI-integrated teaching practices conducted within WP3 of the ELITE-AI project demonstrates the significant potential of AI to enrich language education when used in a structured and pedagogically informed way. Across diverse institutional and disciplinary contexts, a set of coherent patterns has emerged, highlighting both the opportunities and the challenges associated with AI integration. The findings show that AI can support a wide range of learning processes, including writing, reading, speaking, and vocabulary development. However, its most meaningful contribution lies not in automating these processes, but in enhancing them through interaction and reflection. Activities that incorporate revision or transformation of AI-generated content consistently led to deeper engagement and more effective learning outcomes.

At the same time, the analysis underscores the importance of critical engagement with AI. Students must be encouraged to evaluate AI outputs, particularly in contexts where accuracy is essential. This shift towards critical AI literacy represents a key development, aligning language learning with broader educational goals related to digital competence and responsible technology use. The role of the teacher remains central in this process. Far from being replaced, educators become designers of learning experiences, facilitators of interaction, mediators of knowledge and guides in the interpretation of AI-generated content. Their ability to structure tasks and foster critical thinking is essential for ensuring that AI contributes positively to learning.

Overall, the results suggest that the integration of AI into language curricula should be approached not as a technological innovation alone, but as a pedagogical transformation. When thoughtfully implemented, AI has the potential to support more authentic and learner-centred forms of language education, preparing students for the communicative and digital demands of contemporary academic and professional environments.

CHAPTER SEVEN

Applied Guidance for Researchers, Students, Teachers, Evaluators, and Policy Makers. Manuals for Stakeholders

7.1 Introduction

The proposal supporting the ELITE-AI project identified three primary target audiences for the applied outputs of Work Packages 2, 3 and 4: researchers, teachers and students, and policy makers. Following a two-year plan of research and six months of classroom implementation and data collection, exhaustively introduced in the chapters above, the consortium translated these objectives into a series of practical resources in the form of five manuals: *A Manual of Good Practices for Researchers* (ELITE-AI Consortium, 2026a), *Integrating AI Support in ESP Classes: A Manual of Good Practices for Teachers and Students* (ELITE-AI Consortium, 2026b), *A Manual of Good Practices for Policy Makers* (ELITE-AI Consortium, 2026c), *A Quality Assurance Manual* (ELITE-AI Consortium, 2026d), and *An Ethical Provisions Manual* (ELITE-AI Consortium, 2026e).

Each manual draws on a shared evidence base rather than an independent one. The firsthand classroom-facing pair rests most directly on the trial itself: 1,870 individual responses collected across three parallel observation databases, 1,522 in-class student self-assessments, 221 at-home student self-assessments, and 127 teacher observation sheets, gathered across the four partner institutions over two semesters (ELITE-AI Consortium, 2026b), together with the four presentations delivered at the *Workshop Integrating AI Support in ESP Classes* on 7 May 2026 (Gugić, 2026; Li Destri, 2026; Kompara Lukančič, 2026b; Nechifor, 2026a). The *Quality*

Assurance Manual works from a wider perspective of the same trial, considering the integration of over 50 AI-integrated lessons across five courses, four-year groups, and ten distinct AI platforms at the four partner institutions (ELITE-AI Consortium, 2026d). The *Ethical Provisions Manual* (ELITE-AI Consortium, 2026e) adds a second, largely external evidence base, from, on the one hand, the ethical frameworks published by UNESCO (Miao & Holmes, 2023; Miao & Shiohira, 2024; Miao & Cukurova, 2024), the European Commission (2026) [the European Union's *AI Act* and *General Data Protection Regulation* – Regulation (EU) 2024/1689; Regulation (EU) 2016/679)], and the research-ethics guidance of the British Educational Research Association (2024), to a second internal source, on the other hand, the *Stakeholder Workshop for Policy Makers* held on 17 June 2026 (ELITE-AI Consortium, 2026e, 2026f).

This chapter does not treat the five manuals one at a time, as the preceding chapters have already described the template architecture, the theoretical grounding of WP3's observation sheets, as well as their implementation and data collection in detail. Instead, it draws on all five documents together and organises the discussion around the themes that recur across them, regardless of which manual first raised a given point. It also appeals, where relevant, to the lesson-planning chapter of the consortium's earlier monograph, *Empowering Specialised Language Acquisition with Integrated AI* (Blažević et al., 2026), and on independent peer-reviewed analyses of the same evidence base published separately by members of the consortium (Nechifor, 2025a, 2025b, 2026b; Nechifor, Turculeț & Vasilescu, 2025). Section 2 sets out the principle that governs tool selection across all five documents, which draws on the fact that a tool's pedagogical function must precede its adoption. Section 3 describes the observation-sheet and lesson-plan

architecture through which that principle was operationalised as a research and teaching-learning instrument. Section 4 reports the empirical findings the trial produced across the four language skills. Section 5 discusses how responsibility for AI use shifted between teacher and learner over the course of a semester, and how feedback and assessment were redesigned to track that shift. Section 6 turns to disclosure, institutional governance, and the ethical questions the trial and the policy makers' workshop raised, while Section 7 concludes by drawing the four strands together and noting what remains open for the professional development work that future actants involved in decision-making can consider.

7.2 The Pedagogical-Function Principle across the Five Manuals

Needs-based approaches to English for Specific Purposes (ESP) have long held that a course, and by extension any material or tool introduced into it, should be justified by the specific communicative requirements of a defined learner population rather than by the material's general availability or novelty (Hutchinson & Waters, 1987). *The Manual of Good Practices for Researchers* (ELITE-AI Consortium, 2026a) restates this principle for the specific case of AI tools, as its first orientation principle: the pedagogical function of a tool must be established, at a defined lesson stage and for a defined skill and target learner population, before the tool is selected or recommended. A tool adopted because it is free or widely used is not equivalent to a tool chosen for a specific instructional purpose. Similarly, a research question such as whether AI helps ESP students in general is too broad to answer in a single project. The question should specify what the tool is used for, at what stage, and which skill is being investigated. The

same principle recurs, restated for a classroom audience, in *The Manual of Good Practices for Teachers and Students* (ELITE-AI Consortium, 2026b). Li Destri (2026), presenting at the 7 May workshop, formulated it as a distinction between AI as a conversation partner and AI as a substitute teacher, arguing that the technology's usefulness in ESP contexts follows directly from its supplementary role, providing feedback on a draft, suggesting domain-specific collocations, generating listening material on a specialised topic, while the instructor retains responsibility for the instructional frame (ELITE-AI Consortium, 2026b). The same distinction structures *The Quality Assurance Manual*'s account of teacher roles across a semester (ELITE-AI Consortium, 2026d), discussed in Section 5 below, and the manual's own account of general-purpose AI performance across the four partner institutions supports it empirically: general-purpose tools performed reliably on grammar checking and text structuring but underperformed on specialised lexis in domain after domain, a gap Kompara Lukančič's (2026a) work on the *Krajšavar* abbreviation-recognition tool at the University of Maribor demonstrates directly, since *ChatGPT* and *Copilot* both missed or incorrectly expanded criminal-justice and security terminology that a domain-specific tool extracted correctly. Fluency, in other words, is not evidence of domain accuracy, and a tool's output requires the same pedagogical scrutiny regardless of how competent the surface text appears.

The Ethical Provisions Manual restates the principle again, this time at institutional scale and in the register of governance rather than pedagogy. At *The Stakeholder Workshop for Policy Makers* on 17 June 2026, participants were asked, in a live poll, which uses of AI in education concerned them most; 98% of votes named students submitting AI-

generated text without acknowledging it, well ahead of AI-assisted grading by teachers or AI-assisted research planning (ELITE-AI Consortium, 2026e, 2026f). The near-unanimity of that result indicates that what troubles policy makers is undisclosed generation rather than AI use as such, which is the same underlying distinction the researchers' and classroom manuals draw, applied to a question of accountability rather than one of pedagogical design: what determines whether an instance of AI use is justifiable is not whether AI was involved but whether the decision to involve it and the disclosure of that involvement were made explicitly. Read together, the three manuals converge on a single governing claim, phrased differently for a methodological, a pedagogical, and a governance audience, and the remaining sections of this chapter trace how that claim was operationalised, tested, and, in the case of disclosure, found only partially achieved.

Independent evidence from outside the five manuals supports treating tool choice as an empirical question rather than a matter of convenience. Nechifor, Turculeț and Vasilescu (2025) evaluated around 200 mobility-domain ESP scenarios generated by five platforms, *ChatGPT, Claude, Copilot, DeepSeek,* and *Gemini*, against five criteria derived from Hutchinson and Waters' (1987) needs-based ESP framework, Byram's (1997) model of intercultural competence, and Thomas's (1983) pragma-linguistic distinctions, as already discussed in Chapter 5 above: pedagogical authenticity, linguistic appropriateness, cultural sensitivity, discourse coherence, and pragmatic authenticity. Scores varied meaningfully by platform and by criterion, with cultural sensitivity averaging 4.55 and pedagogical value 4.16 on the study's scale, while pragmatic authenticity averaged a comparatively modest 4.03. A validation

exercise, which compared automated ratings generated by Llama 3.1 against human specialist judgement, found only a moderate, non-significant correlation ($\rho = 0.40$, $p = .51$), thus generating two implications that followed for the argument developed in this section. First, general-purpose platforms are not interchangeable even on tasks, such as generating a culturally appropriate mobility scenario, where they might be assumed to perform uniformly well, which reinforces the manuals' insistence that a tool must be evaluated against the specific function it is meant to serve. Second, automated evaluation of AI output against pedagogical criteria is itself only partially reliable, which argues for the kind of human critical evaluation discussed in Section 5.3 rather than for delegating that evaluation to a second AI system.

7.3 Evidence-Based Instrument Design: Observation Sheets and the WP3 Architecture

7.3.1 The Observation Sheet as Research Instrument

The observation sheet is the instrument through which the pedagogical-function principle discussed in Section 2 was made empirically tractable. *The Manual of Good Practices for Researchers* (ELITE-AI Consortium, 2026a) specifies that each sheet should be anchored to a single instructional stage and skill rather than built to cover an entire lesson, on the grounds that a sheet attempting to record everything tends to record nothing with sufficient precision to support analysis. Self-report and behavioural observation are treated as complementary rather than interchangeable measures, and where the two diverge, the divergence itself is read as a finding rather than a problem requiring resolution before analysis can proceed.

The Quality Assurance Manual's own observation sheet operationalises this guidance across six dimensions: communication with the AI, prompt effectiveness, interpretation of output, critical evaluation, target language use, and degree of AI use (ELITE-AI Consortium, 2026d). Participation is tracked across three degrees, adapted from Holec's (1981) account of learner autonomy, i.e., active by volition, active by request, and passive, allowing the instrument to register not only whether AI was used but under what conditions of learner initiative. *The Manual of Good Practices for Teachers and Students* applies a parallel but more classroom-oriented design across three databases distinguishing teacher observation, in-class student self-assessment, and at-home student self-assessment, on the reasoning that a teacher's account of a session, a student's account of the same session, and a student's account of independent use outside the classroom answer different questions and should not rely on a single instrument (ELITE-AI Consortium, 2026b).

7.3.2 Lesson Plans as Research and Teaching Outputs

The most concrete methodological contribution of *The Manual of Good Practices for Researchers* is its treatment of lesson plans as outputs in their own right rather than as incidental teaching materials. Twenty lesson plans, built for B2 to C1, as detailed upon in Chapter 3 above, learners across four ESP domains, follow a five-stage design, according to classical unfolding of a lesson: lead-in, teaching, practice, feedback, and homework, that the manual presents as organising instruction and data collection simultaneously (ELITE-AI Consortium, 2026a). A lesson plan is described as a research-grade when a researcher at a different institution could replicate the instructional sequence from the written record alone, without requiring clarification from the original designer, and this standard

underlies the project's stated progression: from lesson plan to observation sheet to publication, the consortium having committed to six peer-reviewed articles across the three classroom-trial activities as a project deliverable on this basis and having successfully exceeded the envisioned number, due to the abundance of empirical data gathered.

Applied to the trial data reported in *The Manual of Good Practices for Teachers and Students*, the five-stage architecture shows uneven returns across its stages. Thus, lead-in tasks produced the strongest engagement of any stage, based on AI-generated discussion prompts or authentic domain texts, while the feedback stages required the most careful mediation, since in 79 of the 127 observed teacher sessions students used AI in isolation rather than in dialogue with the teacher or with peers (ELITE-AI Consortium, 2026b). This latter pattern is treated as a risk only insofar as it prevents students from connecting a tool's output back to the learning objective of the session, not as an undesirable outcome in itself.

Across the full set of teacher observation sessions, motivation was rated as developing in 82 cases and high in 43; language learning was judged enhanced in 58 sessions and merely maintained in 64; and specialised vocabulary improved in 70 sessions and was still developing in 57 (ELITE-AI Consortium, 2026b). Gugić and Blažević's (2026) experimental data from the University of Pula add an important qualification: AI use produced no statistically significant difference in final test scores between the experimental and control groups when individual sessions lasted only 45 minutes, which suggests that the gains recorded in the observation data, particularly in vocabulary, depend on cumulative exposure across a semester rather than on isolated contact with a tool.

A single lesson plan from the project's own repository, chosen at random out of the 20 produced by the researchers as starting points of any classroom practice empowered by AI tools, illustrates how the five stages were populated in practice. Built for a second-year, B2/C1 cohort of 35 students in a 100-minute session on criminal justice and security terminology (Blažević et al., 2026), the lead-in stage asked students to brainstorm known abbreviations individually before pooling results through *Mentimeter*, an AI tool that clusters submitted terms into a word map; the teaching stage introduced the *Krajšavar* algorithm for recognising abbreviations and their expansions, applied directly to source texts such as official notes; the practice stage had students apply *Krajšavar* to a corpus of authentic texts, extract abbreviation-expansion pairs, and cross-check them against native-language equivalents using *DeepL*; the feedback stage verified translation and spelling accuracy using *DeepL* and *Grammarly* in combination; and the homework stage asked students to compile their own 25-entry database of field-specific abbreviations using *SketchEngine*. No single tool appears at more than one stage performing the same function, which is consistent with the basic orientation principle, as stated in Section 2: each application of AI in the sequence answers a distinct instructional question rather than serving as a general-purpose substitute for the others.

7.4 Empirical Findings across Language Skills

Amassed across the 1,522 in-class responses, 66.6% of students reported that AI had made learning easier and 13.3% reported that it had made learning harder; in the 221 at-home responses, where use was less directly supervised, 72% found the tools helpful and 80% felt they had used them appropriately across the semester (ELITE-AI Consortium, 2026b). Gugić (2026) described the mood these figures reflect as one in which teachers

and students are simultaneously apprehensive about AI and drawn to it, and the data support a correspondingly qualified reading: the benefit recorded is conditional rather than automatic, and its distribution differs noticeably by skill.

A closer analysis of the same underlying dataset complicates the collective satisfaction figures further. Nechifor (2026b), combining the *Student self-assessments* with the *Teacher observation sheets*, documents what she terms an 'integration paradox': reported satisfaction and actual learning behaviour do not follow the same pattern, and a student's stated confidence in a tool is not a reliable predictor of whether that student engaged with the tool's output critically or productively. This finding sharpens rather than contradicts the picture given by the collective figures above, since it is consistent with the writing-skill result reported in Section 7.4.1, where the decisive factor was not whether a student reported the tool as helpful but whether the student could account for the changes an AI suggestion had produced. Satisfaction, according to this reading, is a poor criterion against which successful use of AI should be judged, i.e., the pedagogical-function principle set out in Section 7.2, and an evaluator who relies on self-reported satisfaction alone risks certifying integrations that look successful without having verified that they are.

7.4.1 Writing and Professional Communication

Writing produced the most detailed evidence of any skill in the trial. Li Destri's (2026) first-year tourism students at the University of Udine used *Microsoft Copilot* across five assignments, moving from a professional email draft corrected for mistakes, through a cover letter refined for grammar and vocabulary, to a final piece revised while explicitly

preserving the student's own voice (ELITE-AI Consortium, 2026b). The decisive variable in this sequence was not which stage a student had reached but whether the student could explain, sentence by sentence, why a given AI suggestion had been accepted or rejected; students who submitted AI-refined text without this reflective step showed no measurable gain in writing competence, regardless of how polished the resulting text appeared. This finding is consistent with the broader claim, developed in Section 7.2, that the pedagogical value of a tool depends on the decision surrounding its use rather than on the tool's output alone.

7.4.2 Vocabulary and Domain Terminology

Specialised vocabulary showed the most consistent gain of any skill across the four institutions, improving in 70 of 127 observed teacher sessions (ELITE-AI Consortium, 2026b). The clearest domain-specific evidence comes from the work on the *Krajšavar* algorithm at the University of Maribor. Developed originally for Slovene, extended to English texts in criminal justice and security (Kompara Lukančič & Smajla, 2025) and then tested on Italian tourism texts alongside *Copilot* (Kompara Lukančič, 2026a), the algorithm extracted terminology that general-purpose tools missed or expanded incorrectly. The implication drawn in both *The Manual of Good Practices for Researchers* and *The Manual of Good Practices for Teachers and Students* is that AI-generated vocabulary resources are most reliable when drawn from authentic domain texts rather than from a general-purpose model's default output, and that fluency in a generated definition or gloss is not, on its own, evidence of its accuracy.

7.4.3 Listening and Speaking

Listening and Speaking returned the most equivocal results of the four skills. Conversational and pronunciation applications, including *Gliglish* and *SmallTalk2Me*, appeared throughout the at-home self-assessment data, but their contribution to measurable gains proved harder to isolate than the contribution of AI to writing or vocabulary (ELITE-AI Consortium, 2026b), a pattern consistent with Gugić and Blažević's (2026) null result on the Pula Listening component of the final test discussed in Section 7.3.2. *The Manual of Good Practices for Teachers and Students*' reading is not that AI fails to support speaking practice but that a 45-minute session is too short a window in which to observe the effect, and that unsupervised, lower-anxiety practice at home may be a more productive setting for it. *ChatGPT* was the most frequently named primary tool across the at-home data, used by 51 respondents, followed by *Gemini* in combination with *ChatGPT* and by *Perplexity*. Speaking-specific applications appeared in individual responses without becoming widespread, indicating a gap between availability and actual use that explicit teacher recommendation could plausibly close (ELITE-AI Consortium, 2026b).

7.5 Autonomy, Roles and the Progression of Responsibility

7.5.1 Teacher Roles across a Semester

The Quality Assurance Manual identifies three teacher roles that recur across a semester of AI-integrated instruction, and treats the movement between them, rather than any single role in isolation, as the primary indicator of programme quality (ELITE-AI Consortium, 2026d). Early

sessions position the instructor as a ‘curator’, assembling source material and presenting AI-generated output as structured input for students to work from. Middle sessions position the instructor as a ‘scaffolder’, modelling a productive interaction before students attempt it unassisted, a movement the manual reads through Vygotsky’s (1978) concept of the ‘zone of proximal development.’ Later sessions position the instructor as an ‘evaluator’, stepping back to allow students to prompt independently while directing attention to the critical engagement that follows. A course in which students use AI in the same way in the final session as in the first has not, on this account, achieved the autonomy development the project treats as a primary outcome, independent of other measures of course quality.

7.5.2 Learner Autonomy Profiles

The Manual of Good Practices for Teachers and Students distinguishes three profiles of AI use among learners: ‘directed use’, where students engage with a tool because the instructor requires it; ‘proactive use’, where students seek out AI support on their own initiative; and ‘avoidance’, where students decline to use AI at all. In the in-class data, 54.5% of students reported directed use, 35.8% reported proactive use, and 9.7% reported avoidance; by the end of the semester, the at-home data show a shift toward the proactive category relative to the earlier, more directed in-class figures (ELITE-AI Consortium, 2026b). This pattern is consistent with Hashim’s (2018) characterisation of contemporary learners as comfortable with new technology provided it is introduced with sufficient guidance to build confidence rather than confusion, and it supports the manual’s broader

claim that the 37%[5] of in-class respondents who struggled to connect AI use to their own learning did so not because AI was unhelpful but because the connection had not yet been made explicit to them by either the instructor or the tool itself.

7.5.3 Feedback and Critical Evaluation

Feedback quality in *The Quality Assurance Manual* is assessed against Nicol and Macfarlane-Dick's (2006) seven principles of good feedback practice, and the manual's clearest illustration of the principle in operation involves a single UNITBV session in which 16 students submitted transcribed role-play exchanges to *ChatGPT* and received individualised written commentary on register and functional coherence simultaneously, a volume of personalised feedback no single instructor could have produced single-handedly within the available time (ELITE-AI Consortium, 2026d). The tool did not substitute for the instructor's judgement about what the commentary should evaluate but it extended a capacity the instructor already possessed.

Assessment in *The Manual of Good Practices for Teachers and Students* evaluates four dimensions of an AI-assisted assignment: *the linguistic quality of the finished text, the appropriateness of the AI use itself, the quality of the prompts that produced it,* and, weighted most heavily, *the student's demonstrated ability to evaluate the AI's output critically*

[5] These figures were extracted from the in-class self-assessment database while data collection was still under completion. They therefore differ from the percentages reported by Nechifor (2026b) and cited in Chapter 5, Section 5.3 (66.3% directed use and 32.7% independent initiative), which were calculated once data collection had been completed. Both sets of figures are reproduced as they appear in their respective sources, and the difference between them reflects the stage of data collection at which each was obtained.

(ELITE-AI Consortium, 2026b). The self-assessment data indicate that students already practise a version of this independently: the most commonly cited reason for verifying an AI-generated answer was familiarity with the concept of *hallucination*, the tendency of a language model to produce a plausible but fabricated statement, and students who already knew the term verified output more carefully than those who did not. This finding is treated in the manual as evidence that critical-evaluation competence is teachable through explicit naming of the relevant concept rather than something that develops only through incidental exposure, a point returned to in the conclusion of this chapter.

7.5.4 Connecting Grammatical and Pragmatic Competence

The critical-evaluation competence discussed above sits inside a broader theoretical question about what AI-assisted ESP instruction is meant to develop in the first place. Nechifor (2025b) proposes a comprehensive framework connecting grammar development and communicative competence in AI-enhanced ESP acquisition, arguing that GenAI tools are most defensible pedagogically when they are used to strengthen the link between grammatical accuracy and pragmatic appropriateness in professional contexts, rather than when they are used to produce grammatically correct output in isolation from the communicative situation it is meant to serve. This framework gives the autonomy progression described in Section 7.5.1 a linguistic rationale beyond the general claim that autonomy is desirable: a learner who moves from directed to proactive AI use without also being able to judge whether a generated utterance is pragmatically appropriate to a given professional register has developed operational independence without developing the specific competence ESP instruction is meant to target. The reflective step identified as decisive for

writing gains in Section 7.4.1, in which a student explains why an AI suggestion was accepted or rejected, is one concrete mechanism through which that link between grammar and pragmatics can be made explicit during instruction rather than left implicit.

7.6 Ethics, Disclosure and Institutional Governance

7.6.1 Disclosure Norms in the Classroom

Nechifor's (2026a) contribution to the 7 May workshop identified an asymmetry that the following part of this section develops in more detail: academic publishing has developed AI-disclosure requirements more quickly and more explicitly than classroom practice has, with the consequence that a student trained only on classroom norms may reach a first submission to a publisher without having been shown that the two settings apply different, and sometimes contradictory, conventions. The self-assessment data quantify the resulting uncertainty, as when asked whether they had integrated AI-generated content ethically into their assignments, only 103 of 221 at-home respondents answered 'yes'; 110 gave an answer closer to 'more or less', and eight answered 'no' (ELITE-AI Consortium, 2026b). This distribution is read in the manual not as evidence of widespread dishonesty but as evidence of an unresolved question that individual students cannot be expected to settle for themselves, consistent with the WP3 classroom findings reported in *The Ethical Provisions Manual*: disclosure norms emerged spontaneously in the earliest sessions of the trial, ahead of any formal institutional guidance in every partner country (ELITE-AI Consortium, 2026e).

7.6.2 Publisher Requirements

The Manual of Good Practices for Teachers and Students sets classroom practice against the requirements of four major academic publishers relevant to the ELITE-AI fields. Elsevier permits AI-assisted writing conditional on a named declaration and independent verification of every AI-suggested reference. Oxford University Press requires the editor's written approval before any AI-generated content is submitted. Cambridge University Press, in October 2024, wrote to 20,000 of its own authors requesting explicit consent before licensing their published work for model training. Peter Lang, the strictest of the four, does not accept a manuscript that has been written or substantially generated by AI in any part, and permits the technology only for language checking and administrative tasks (ELITE-AI Consortium, 2026b). Three principles recur across all four policies despite their differences: accountability for AI-assisted work cannot be delegated to the tool that helped produce it; some form of disclosure is required even where its precise scope varies by publisher; and confidentiality during peer review is treated as absolute, a convention the manual notes is rarely taught explicitly in classrooms and is, as a consequence, frequently violated by students unaware that it exists.

7.6.3 Policy Maker Perspectives and Institutional Precedents

The Ethical Provisions Manual's governance chapters draw on a *Stakeholder Workshop for Policy Makers* held online on 17 June 2026, which drew approximately 95 participants from at least a dozen institutions and countries (ELITE-AI Consortium, 2026e, 2026f). Beyond the finding on undisclosed generation reported in Section 7.2, a separate poll found 86% of respondents in favour of mandatory AI-ethics training for staff, and

an open-text question on institutional barriers identified a lack of AI literacy among staff as the most frequently endorsed obstacle, ahead of the absence of any regulation at all. Contributions from the floor, from participants affiliated with the University of Santiago de Compostela, the University of Udine, a Ghanaian institution, and a Slovene-language programme at a Chinese university, illustrated a wide spread of institutional starting points, from the complete absence of AI-specific regulation, through an approved policy unread by most of the institution, to mandated, top-down adoption already reshaping curriculum.

A comparison of five universities with published AI policies, including Stanford University, the University of Oxford, the University of Cambridge, and KU Leuven, identifies three points of convergence that hold regardless of national legal systems. Disclosure obligations are now close to universal but are calibrated to the risk associated with a specific use rather than applied uniformly to AI use as such. Authority over policy is decentralised to the course or discipline level in nearly every case examined, since a single fixed institutional rule cannot appropriately cover either the extent of diversity and specificity particular areas require or, even more properly, the range of subjects a university teaches. And every policy reviewed, including the University of Cambridge's, which was revised twice within a year of first publication, describes itself as provisional (ELITE-AI Consortium, 2026e). A post-workshop questionnaire completed by 32 policy makers corroborates the poll results, and the manual's resulting professional-development recommendations follow directly from what participants reported: training in advance of enforcement, a named institutional contact responsible for keeping any policy current, active dissemination of whatever policy is approved, and peer observation on the

WP3 model in place of AI-detection software, which the manual treats as unreliable. Its closing checklist consolidates these recommendations under nine headings, from data privacy and informed consent through pedagogical integrity and equity to continuous evaluation and professional development, intended for direct institutional use rather than sequential reading.

7.6.4 The Equity Dimension of Digitalisation

The governance findings reported above concern how institutions regulate AI once access to it is assumed. A separate and logically prior question, raised by Nechifor (2025a) in an analysis of what she terms 'the *globitalised* discourse of education', is whether that access can be assumed at all. Digitalisation of language acquisition carries genuine advantages, such as individualised access to instruction regardless of geographic location, adaptability to diverse learning styles, and reduced dependence on any single institution's physical resources, but it carries corresponding risks that the governance frameworks discussed in this section do not, on their own, address: unequal connectivity, reduced opportunities for direct human interaction, and the isolation of learners who lack reliable access to the infrastructure a digitalised curriculum assumes. This dimension is consistent with the access caution raised independently by *The Manual of Good Practices for Researchers* in Section 7.2 and restated by *The Ethical Provisions Manual* as an equity criterion in its consolidated checklist; the present chapter's contribution is to note that the caution applies to the project's own applied outputs as much as to the institutions they address, since a manual, however well designed, assumes a reader with the infrastructure to act on it.

7.7 Conclusion: Toward Coherent Guidance across Audiences

The five manuals discussed in this chapter address distinct audiences and were written to distinct conventions, a methodological guide for researchers, a practical guide for teachers and students, an audit instrument for evaluators, and a governance brief for policy makers, but the evidence reviewed in Sections 7.2 through 7.6 indicates that they rest on a single underlying claim rather than four unrelated sets of recommendations. What determines whether an instance of AI use in ESP instruction is pedagogically, methodologically, or ethically defensible is not the presence or absence of the technology but the explicitness of the decision surrounding its use: which instructional function it was meant to serve, whether that function was disclosed to whoever is entitled to know about it, and whether the person using the tool remained able to evaluate its output critically rather than accepting it by default.

Two further points follow from the evidence reviewed here and are worth explicit statement. First, none of the five manuals, nor any of the five institutional policies reviewed in Section 7.6.3, presents itself as a finished statement; all describe their own recommendations as provisional and subject to revision as the evidence base grows, a posture the consortium's own practice, revising *The Ethical Provisions Manual* for *The Policy Makers' Manual*'s production following internal review in August 2026, has already demonstrated. Second, the capacity most consistently identified as under-developed across the evidence, named as either 'critical evaluation', '*hallucination*-literacy', 'verification', or 'disclosure competence' depending on which manual is consulted, is treated in each case as something a student or teacher will eventually acquire, typically

through a single explicit session, a checklist item, or incidental exposure, rather than through any structured instructional sequence. The finding reported in Section 7.5.3, that students who already knew the term *hallucination* verified AI output more carefully than students who did not, indicates that this competence responds to direct instruction; the disclosure data reported in Section 7.6.1 indicate that, at present, such instruction is inconsistent. Addressing that inconsistency, rather than producing further guidance documents, is likely to be the more consequential task for the professional development work the project undertook when it aimed for a lasting impact.

All five manuals, together with the WP2 lesson-plan repository, the observation-sheet templates, and the peer-reviewed articles the consortium committed to as a project deliverable, are freely available at www.elite-ai.unitbv.ro.

Conclusions

This monograph began with a question about language before it turned to any question about technology. Chapter 1 described language, following Grosman (2008), as a home that its speakers adapt to their needs throughout their lives, and traced the development of Language for Specific Purposes (LSP) and English for Specific Purposes (ESP) as responses to the needs of people who use a language in their studies and work. Chapter 2 then introduced artificial intelligence (AI) against this background as a set of tools whose capacity to personalise practice and give immediate feedback comes with real risks for privacy and for students' independent thinking. The chapters that followed reported what happened when four universities, in Brașov, Maribor, Udine and Pula, introduced those tools into their ESP courses under a common design, methodology and research protocol. This concluding chapter underlines what that work allows the authors to claim and considers where it points for ESP teaching after the project.

The Decision before the Tool

Across the chapters, written for different purposes and from different evidence, one claim recurs often enough to be treated as the monograph's central finding: the value of an AI tool in ESP teaching depends on the decisions made around it, about the communicative need it is meant to serve, the lesson stage at which it is used, the person who decides to use it and the way its output is checked. The *Manual of Good Practices for Researchers* states this as its first orientation principle, according to which a tool's pedagogical function must be established before the tool is adopted (ELITE-AI Consortium, 2026a). Chapter 3 showed that the *AI Tools*

Repository and the *Interactive Lesson Plan* include the same requirement into their structure, since no tool could enter the repository without a skill and a stage, and no activity could enter a lesson plan unless its content focus, lesson stage and tool were specified together (ELITE-AI Consortium, 2025c, 2025d). In Mishra and Koehler's terms (2006), each completed cell of the template is a small-worked example of technological pedagogical content knowledge (TPCK). Chapter 6 reached the same conclusion from classroom practice, finding that the most successful activities were those in which AI had a defined purpose within the task and was accompanied by critical evaluation and discussion. Chapter 7 found it once more in the field of governance, where 98% of the votes expressed at the *Stakeholder Workshop for Policy Makers* named the submission of unacknowledged AI-generated text as the most worrying use of AI in education, well ahead of AI-assisted grading or research planning (ELITE-AI Consortium, 2026e, 2026f).

The principle extends nowadays the needs-based logic that Hutchinson and Waters (1987) placed at the centre of ESP from the choice of content and method to the choice of tools. Seen in this light, integrating AI into ESP calls for the consistent application of a criterion ESP teachers already know, which is a practical advantage in a market of tools that changes faster than any syllabus. It also answers, at the level of design, the weakness Stockwell (2007) identified in CALL research, where technology was often chosen for convenience and reported without an explicit pedagogical rationale. Whether the answer holds in practice depends on whether the teachers who open a lesson plan share the judgement that produced it, a question Chapter 3 left open and one to which this chapter returns below.

Perception, Use and Achievement

The second conclusion concerns the nature of the evidence, with chapter 5 having insisted on the idea that perceived usefulness, patterns of use and measured achievement are distinct kinds of data, the project's results showing why that distinction must be maintained. In this way, students' judgements of AI-supported activities were largely favourable, a fact made visible in the analysis reported by Nechifor (2026b), which recorded from the project data that 91.1% of student assessments described the integration of AI as helpful, and from the *Manual of Good Practices for Teachers and Students*, where 66.6% of the in-class self-assessments summarised stated that AI had made learning easier (ELITE-AI Consortium, 2026b). The same data show, however, that a favourable judgement and a productive use of the tool frequently diverged. Nechifor (2026b) found that 28.7% of students accepted generated responses without questioning their accuracy and that many used AI for isolated tasks without linking them to the wider learning process, a mismatch she referred to as 'the integration paradox'. The achievement data add a further qualification, since in the precise comparison conducted by Gugić and Blažević (2026) the experimental and control groups did not differ significantly in final achievement once initial knowledge was taken into account.

Nevertheless, neither set of findings amounts to a verdict on AI in ESP. The comparison by Gugić and Blažević (2026) involved 40 students, 12 teaching hours per circumstance and an experimental situation that also gave students more autonomy than the control context, so the effect of the tool cannot be separated from the effect of the approach that surrounded it. The favourable perceptions say a good deal about how students

experienced the activities and much less about what the activities changed. Read together, the findings support the position Chapter 7 drew from Nechifor's (2026b) analysis, namely that satisfaction is a weak criterion of successful integration and that an evaluator who relies on it alone may approve uses of AI that appear successful without having established that they are. For future evaluations, within and beyond this project, the practical consequence is that data on perception, behaviour and achievement should be collected together and reported separately, as the ELITE-AI protocol set out to do.

Skills, Domains and the Limits of Fluency

The evidence was also uneven across language skills, and the unevenness is itself informative. Within the project, writing produced the most detailed evidence and specialised vocabulary the most consistent gains, with vocabulary judged to have improved in 70 of the 127 observed teacher sessions (ELITE-AI Consortium, 2026b). Listening and speaking produced the least conclusive results. As Chapter 3 noted, this pattern runs close to the opposite of the meta-analysis by Wu (2024), which found AI interventions most effective for listening and speaking. Goh and Aryadoust (2025) suggest a plausible reconciliation, observing that much of the evidence for gains in these skills rests on learners' engagement and their sense that a conversational system feels humanlike, and much less on measured linguistic development. Because the ELITE-AI instruments recorded engagement and outcomes separately, the two could diverge in the project's data. Session length is a further factor, since a 45-minute class offers a narrow window in which to observe change in oral skills, and the

suggestion in the *Manual of Good Practices for Teachers and Students* that lower-anxiety practice at home may suit these skills better deserves to be tested directly (ELITE-AI Consortium, 2026b).

The writing results carry a lesson of their own. In Li Destri's (2026) tourism classes at the University of Udine, the students who gained more writing knowledge were those who could explain why each AI suggestion had been accepted or rejected, whereas polished text submitted without that reflection brought no measurable improvement in writing competence (ELITE-AI Consortium, 2026b). Chapter 6 reached a similar conclusion, identifying the revision stage of AI-assisted writing as the point at which learning actually takes place, and both findings agree with Nechifor's (2025b) argument that GenAI is most justifiable in ESP when it strengthens the link between grammatical accuracy and pragmatic appropriateness in a professional situation. However, Shi et al. (2025) add a caution the project's instruments were not designed to detect. Students who wrote with ChatGPT outperformed a group using automated writing evaluation on writing performance but reported a weaker ideal L2 writing self, which suggests that better texts and a stronger sense of oneself as a writer in the second language do not necessarily develop together.

But most importantly, the domain-related findings have the heaviest consequences over what makes ESP distinctive. The *Quality Assurance Manual* reports that general-purpose tools handled grammar checking and text structuring reliably across the partner institutions but underperformed on specialised lexis (ELITE-AI Consortium, 2026d), and the work on the *Krajšavar* algorithm illustrates the alternative, being a tool built for one precise operation on specialised texts, tested on English criminal justice

and security material and compared with *Copilot* on Italian tourism texts (Kompara Lukančič & Smajla, 2025; Kompara Lukančič, 2026a). Nechifor, Turculeț and Vasilescu (2025) found a related imbalance in learning scenarios generated by five platforms for students preparing for academic mobility. The scenarios scored well on cultural sensitivity (4.55) and pedagogical value (4.16) but lower on discourse coherence (3.92) and linguistic appropriateness (3.88), and the automated evaluation the authors tested agreed only partly with the ratings of human specialists ($\rho = .40$, $p = .51$). For a discipline in which the language to be learned is defined by a professional community, this is the finding with the widest consequences. A generated text can read smoothly and still misuse a term or misjudge a register, and such errors become harder to see as the tools grow more fluent. Consequently, the need for a teacher who knows the field increases accordingly.

Teachers and Learners

Nonetheless, none of the chapters found that AI reduced the need for teachers, while several found exactly the opposite. Chapter 2 concluded from the literature review that AI should complement the language instructor, and the classroom evidence gave that conclusion a more precise content. Chapter 6 described the teacher in AI-supported lessons as a designer of learning experiences, a facilitator, a mediator of knowledge and, in specialised domains, an epistemic guide who helps students separate accurate from inaccurate information. The *Quality Assurance Manual* described a progression across the semester in which the teacher first curates AI output as structured input, then models productive interaction

before students attempt it alone, and finally steps back to evaluate the critical engagement that follows; a course in which students use AI in the final session exactly as they did in the first has not, on this account, developed their autonomy (ELITE-AI Consortium, 2026d). The self-assessment data are consistent with such a progression, considering that of the in-class self-assessments, 54.5% reported 'directed use of AI', 35.8% 'proactive use' and 9.7% 'avoidance,' while the at-home data collected at the end of the semester showed a shift towards 'proactive use' (ELITE-AI Consortium, 2026b).

However, autonomy is only part of the goal, as a learner who moves from directed to independent use of AI without learning to judge whether a generated utterance suits a professional register has acquired operational independence without the competence ESP sets out to develop (Nechifor, 2025b). This explains the weight the chapters place on critical evaluation, and it makes the finding that critical evaluation develops mainly when tasks require it one of the most useful the project produced. Chapter 6 showed that students tended to trust AI output at first and became more sceptical only in activities that built verification into the task, such as comparing an AI-generated summary of a national agency or an international organisation with its official description or checking an AI-drafted Learning Agreement against the institutional template. Chapter 7 reported that students who already knew the term *hallucination* verified AI output more carefully than those who did not (ELITE-AI Consortium, 2026b), which suggests that the competence responds to direct instruction and need not be left to chance.

Critical AI Literacy in the ESP Curriculum

If the ability to evaluate AI output can be taught, it should be taught deliberately, with an equal focus of intention ESP places on its characteristic features such as genre, register and terminology, which, at present moment, is not implemented. Chapter 7 observed that across the project's manuals the competence variously described as 'critical evaluation', 'verification', '*hallucination* literacy' or 'disclosure competence' is expected to emerge from a single session, a checklist item or incidental exposure, but the empirical data show that this comes at a cost: asked whether they had integrated AI-generated content ethically into their assignments, only 103 of the 221 respondents in the at-home self-assessment observation sheets answered 'yes', 110 gave an answer closer to 'more or less', and eight answered 'no' (ELITE-AI Consortium, 2026b). The *Manual of Good Practices for Teachers and Students* reads these answers as uncertainty about rules that individual students cannot be expected to establish and observe for themselves, all the more so because the rules differ between settings: academic publishers range from permitting AI-assisted writing with a declaration to accepting almost no AI involvement in a manuscript, and classroom practice has developed disclosure norms more slowly than publishing has (ELITE-AI Consortium, 2026b; Nechifor, 2026a).

European regulation has moved in a direction that supports this conclusion. *The Digital Omnibus on AI*, which entered into force in July 2026, replaced Article 4 of the AI Act so that providers and deployers of AI systems must now take measures to support the AI literacy of their staff and of others who operate AI systems on their behalf, without being required to

guarantee any specific level of literacy for any individual (Regulation (EU) 2026/1744, Article 1(5)). The regulation's preamble describes the development of AI literacy, beginning in education and training and continuing throughout life, as crucial, and adds that AI literacy 'should be a strategic priority, regardless of regulatory obligations and potential sanctions' (Regulation (EU) 2026/1744, Preamble, para. 8). The legal obligation concerns the staff of benefactors and operators, but this emphasis on education speaks directly to universities, and ESP courses are well placed to take on part of that work, since they already train students to measure language against the conventions of a professional community.

In practice, such teaching could take the form of a sequence spread across a semester and across the five stages of the ELITE-AI lesson template. Students would learn early in a course to name the typical failures of generative models, *hallucination* among them. Verification against authoritative domain sources would become a regular part of the practice and feedback stages, and students would be asked to account in writing for each AI suggestion they accept or reject. Disclosure would be a routine element of every assignment, with the conventions of academic publishers introduced before students first submit work beyond the classroom. Assessment would then follow the weighting already proposed in the *Manual of Good Practices for Teachers and Students*, in which a student's demonstrated ability to evaluate AI output counts for more than the polish of the finished text (ELITE-AI Consortium, 2026b).

Limitations

These conclusions rest on evidence with some limits, most of which the preceding chapters have already acknowledged. The first concerns the experimental comparison reported by Gugić and Blažević (2026), which involved small groups and a short intervention, with its experimental situation that differed from the control context in more than the use of AI. Another identifies the fact that much of the evidence consists of teachers' observations of their own classes and students' self-reports, both of which are open to expectancy effects, social desirability and recall bias, as Chapter 4 discussed. Also, the applied research instruments, i.e., the observation sheets, recorded what students said about whether they had checked AI output and also how teachers judged their prompts, but not the prompts themselves, the outputs received or the revisions made. Moreover, the four ESP domains covered by the project represent a small part of the field, and a teacher of medical, engineering or maritime English would need to repeat the domain-specific work each partner did for its own students before the lesson plans could serve another field, even though the universality of the methodology can successfully compensate in this respect. And last, some of the tools changed during the project, having some of those named in the repository replaced or discontinued by the time the first monograph was published due to the rapid pace at which technology advances in the dynamic landscape they belong to (Blažević et al., 2026). In all these respects the project shares the condition of a research area that Costantini, Gallitelli and Li Destri (2026) describe as 'still largely exploratory' and heavily reliant on studies of perception.

Directions for Research and Practice

Several directions follow from these limits, the first of which concerns the research design. Thus, future studies would benefit from longer interventions, delayed post-tests and designs that separate the effect of an AI tool from the greater autonomy that tends to accompany it, for example by giving control groups the same independence without the tool. Also, forthcoming attempts should also record what learners actually do with AI, collecting prompts, outputs and successive drafts with informed consent and in line with the *General Data Protection Regulation* (Regulation (EU) 2016/679) and the research-ethics guidance of the British Educational Research Association (2024), because only records of this kind can show how a learner reaches the decision to accept or reject a suggestion. Moreover, constructs that the project measured broadly, such as motivation, could be studied more specifically, as Shi et al. (2025) did with the ideal L2 writing self. Additionally, the cognitive load of activities that combine several tools within a few minutes, a question Chapter 3 raised with reference to Sweller (1988) and Mayer and Moreno (2003), also remains to be measured in the classroom.

A second direction concerns the resources. Chapter 3 argued that a catalogue organised around named tools ages faster than one organised around functions, and a revised repository could treat its functional categories, among them content generation, feedback, text analysis and interactive engagement, as a stable structure to which replaceable tools are attached. Furthermore, domain-specific resources deserve more attention than they have so far received. The extension of *Krajšavar* from Slovene to English criminal justice texts and Italian tourism texts shows that

instruments designed for authentic specialised material can travel across languages (Kompara Lukančič & Smajla, 2025; Kompara Lukančič, 2026a), and the same approach could support LSP courses in the other languages of the consortium, which Chapter 1 placed alongside English under the broader heading of Language for Specific Purposes.

The third direction concerns the people, and it returns to the question left open earlier in this chapter, whether the teachers who use the project's resources share the judgement that produced them. In this respect, Chapter 7 reported that participants in the policy makers' workshop identified a lack of AI literacy among staff as the most frequently cited institutional barrier, ahead of the absence of regulation, and that 86% favoured mandatory AI-ethics training for staff (ELITE-AI Consortium, 2026e and 2026f). Teacher education is therefore where the project's outputs are most likely to have a lasting effect. Likewise, the lesson plans were written to a standard that would allow any colleague at another institution to reproduce them from the written record alone (ELITE-AI Consortium, 2026a), and peer observation on the WP3 model gives institutions a way of supporting good practice that does not rely on AI-detection software, which the *Ethical Provisions Manual* treats as unreliable (ELITE-AI Consortium, 2026e). Moreover, as Nechifor (2025a) argues, the digitalisation of language learning assumes a level of connectivity and infrastructure that not every learner enjoys, and any wider use of the project's resources will need to take account of students and institutions with fewer means. Last but definitely not least, equity belongs to this direction as well.

A Closing Perspective

ESP has its origins in the years after the Second World War, when growing numbers of people needed English for specific professional purposes that general courses did not serve (Hutchinson & Waters, 1987; Rahman, 2015), and the discipline has since adapted to new professions and new media while keeping its central question about what a defined group of learners needs. Generative AI adds a new element to the answer. Students preparing for careers in criminal justice, tourism, education or participating in international academic exchanges now need, alongside the terminology and genres of their field, the ability to judge language produced by a machine and to say openly when they have used it. As Chapter 6 observed, these students are likely to rely on AI tools in their future work to produce the kinds of texts they practise in class, so a course that teaches such judgement prepares them for the conditions in which they will actually communicate.

The evidence gathered by the ELITE-AI consortium gives a fairly concrete picture of what such a course could look like. Tools would be chosen for what they do at a particular stage of a lesson, and students would be expected to account for the suggestions they accept. Verification against authoritative sources would belong to ordinary practice, and the teacher's knowledge of the language and of the field would set the standard against which machine output is measured. The experience of the four partner universities suggests that much of this can be achieved within existing ESP courses through careful task design, and it also shows how much remains to be learned about the long-term effects of such an approach. Chapter 1 closed with the observation, drawn from Moeller and Catalano (2015), that English as a shared language is a means of more effective interaction between people and of better understanding between cultures. AI may help ESP learners reach that goal sooner, provided that students and teachers keep the judgement about language in their own hands.

References

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*(1), Article 10. https://doi.org/10.1186/s41239-024-00444-7

Ahmed, S. M. A. A., Taha, A. R. A., Hussain, S., & Hayat, A. (2023). Enhancing the teaching and learning of English for specific purposes (ESP) with ChatGPT. *International Journal of Technology and Education Research, 1*(3), 40–49.

Aijun, Y. (2024). On the influence of artificial intelligence on foreign language learning and suggested learning strategies. *International Journal of Education and Humanities, 4*(2), 107–120.

Almutairi, H., Alfaifi, A. A., & Saleem, M. (2026). Writing accuracy: How AI-assisted writing instruction can support EFL undergraduate students. *Information, 17*(2), Article 157. https://doi.org/10.3390/info17020157

Amaria, F. (2025). The impact of AI writing tools on academic writing skills in English learning classroom. *International Journal of Language and Education Research, 7*(3), 159–177.

Amebis. (n.d.). *Krajšavar* [Abbreviation recognition tool]. Retrieved October 7, 2026, from http://peter.amebis.si/krajsave.html

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, Article 87. https://doi.org/10.3389/feduc.2019.00087

Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice, 17*(2), 199–214.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(2), 3–13.

Basturkmen, H., & Elder, C. (2004). The practice of LSP. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 672–694). Blackwell.

Benabdallah, A. (2023). Investigating the future of ESP teaching in the age of artificial intelligence. *Annals of Stefan cel Mare University of Suceava: Philosophy, Social and Human Disciplines Series, 2*, 77–93.

Berns, M. (2009). English as lingua franca and English in Europe. *World Englishes, 28*(2), 192–199.

Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: An exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia, 14*(1), Article 45. https://doi.org/10.1186/s40468-024-00322-z

Blažević, I., Costantini, F., Dimulescu, C., Gallitelli, E., Gugić, J., Kompara Lukančič, M., Li Destri, E., Nechifor, A., & Smajla, T. (2026).

Empowering specialised language acquisition with integrated AI. Cambridge Scholars Publishing.

Boeru, M. (2024). Exploring the use of AI tools in teaching English for specific purposes (ESP). *Scientific Bulletin of Naval Academy, 27*(1).

Borich, G. D. (2016). *Observation skills for effective teaching: Research-based practice* (7th ed.). Routledge.

Boud, D. (1995). *Enhancing learning through self-assessment*. Kogan Page.

Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self-assessment and reflection in professional courses. *Social Work Education, 18*(2), 121–132.

British Educational Research Association. (2024). *Ethical guidelines for educational research* (5th ed.). https://www.bera.ac.uk/wp-content/uploads/2024/04/Ethical-Guidelines-for-Educational-Research-5th-edition.pdf

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367–393). SAGE.

Budiyono, H., Marzuki, Pudjaningsih, W., Prastio, B., & Maulidina, A. (2025). Exploring long-term impact of AI writing tools on independent writing skills: A case study of Indonesian language education students. *International Journal of Information and Education Technology, 15*(5), 1003–1013. https://doi.org/10.18178/ijiet.2025.15.5.2306

Burns, H. L., & Capps, C. G. (1988). Foundations of intelligent tutoring systems: An introduction. In M. C. Polson & J. J. Richardson (Eds.),

Foundations of intelligent tutoring systems (pp. 1–20). Lawrence Erlbaum Associates.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Cambridge University Press & Assessment. (2024). *Uncharted territory: AI and the Cambridge approach for academic book publishing*. https://www.cambridge.org/us/universitypress/about-us/news-and-blogs/uncharted-territory-ai-and-the-cambridge-approach-for-academic-book-publishing

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology, 15*(3), Article 39. https://doi.org/10.1145/3641289

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

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 849–874). Elsevier.

Costantini, F., Gallitelli, E., & Li Destri, E. (2026). Integrating AI in education and in ESP teaching: An overview of recent advances. *Journal of Applied Linguistic and Intercultural Studies, 6*, 88–99.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Dasborough, M. T. (2023). Awe-inspiring advancements in AI: The impact of ChatGPT on the field of organizational behaviour. *Journal of Organizational Behaviour, 44*(2), 177–179. https://doi.org/10.1002/job.2695

Debreli, E., & Ishanova, I. (2019). Foreign language classroom management: Types of student misbehaviour and strategies adapted by the teachers in handling disruptive behaviour. *Cogent Education, 6*(1), Article 1648629. https://doi.org/10.1080/2331186X.2019.1648629

Dewey, M. (2007). English as a lingua franca and globalization: An interconnected perspective. *International Journal of Applied Linguistics, 17*(3), 332–354.

Díaz Santos, G. (2000). Technothrillers and English for science and technology. *English for Specific Purposes, 19*(3), 221–236.

Dou, W. (2024). The application of generative AI technology in ESP courses for civil engineering majors. In *2024 International Conference on Artificial Intelligence and Digital Technology (ICAIDT)* (pp. 144-147). IEEE.

Doughty, C. J., & Long, M. H. (Eds.). (2003). *The handbook of second language acquisition*. Blackwell.

Downes, W. (1998). *Language and society* (2nd ed.). Cambridge University Press.

Du, J., & Alm, A. (2024). The impact of ChatGPT on English for academic purposes (EAP) students' language learning experience: A self-determination theory perspective. *Education Sciences, 14*(7), Article 726. https://doi.org/10.3390/educsci14070726

Dudley-Evans, T. (1997). Five questions for LSP teacher training. In R. Howard & G. Brown (Eds.), *Teacher education for languages for specific purposes* (pp. 58–67). Multilingual Matters.

Dudley-Evans, T. (2001). English for specific purposes. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 131–136). Cambridge University Press.

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157–169.

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies, 10*(2), 121–142.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 282–304.

ELITE-AI Consortium (2025a). Project homepage and results overview. www.elite-ai.unitbv.ro

ELITE-AI Consortium (2025b). *ELITE-AI Literature Review Handbook*. https://shorturl.at/LvfAF

ELITE-AI Consortium (2025c). AI Tools Repository. https://shorturl.at/wLbBr

ELITE-AI Consortium (2025d). Interactive Lesson Plan. https://elite-ai.unitbv.ro/index.php/interactive-lesson-plan/

ELITE-AI Consortium (2026a). *A Manual of Good Practices for Researchers*. Work Package 2, Activity 4. https://shorturl.at/LpcaT

ELITE-AI Consortium. (2026b). *Integrating AI Support in ESP Classes: A Manual of Good Practices for Teachers and Students*. Work Package 3. Activity 4. https://shorturl.at/jR3DT

ELITE-AI Consortium. (2026c). *Manual of Good Practices for Policy Makers*. Work Package 4, Activity 4. https://tinyurl.com/4uu7bdp3

ELITE-AI Consortium. (2026d). *Quality Assurance Manual.* Work Package 4, Activity 1. https://tinyurl.com/2an2p2cj

ELITE-AI Consortium. (2026e). *Ethical Provisions Manual.* Work Package 4, Activity 2. https://tinyurl.com/3undwnz6

ELITE-AI Consortium. (2026f). Slido poll results and evaluation questionnaire, Stakeholder Workshop for Policy Makers, 17 June 2026. https://bit.ly/4AV2fdE

ELITE-AI Consortium. (2026g). Observation Sheets Database. Work Package 3, Activities 1, 3 and 4. https://elite-ai.unitbv.ro/index.php/elite-ai-observation-sheets/

ELITE-AI Consortium. (2026h). Research Articles. Work Package 3. https://elite-ai.unitbv.ro/index.php/research-articles-elite-ai/

Elsevier. (n.d.). *The use of generative AI and AI-assisted technologies in writing for Elsevier*. https://www.elsevier.com/about/policies-and-

standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier

Erito, S. N. P. (2023). Exploring ESP students' perception toward the potential of artificial intelligence to promote students' self-efficacy in English writing skill. *Journal of English Language Learning, 7*(2), 457–464. https://doi.org/10.31949/jell.v7i2.7598

European Commission: Directorate-General for Education, Youth, Sport and Culture. (2026). *2026 guidelines on the ethical use of artificial intelligence and data in teaching and learning for educators*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/f692aa0b-17a7-11f1-8870-01aa75ed71a1/

Fattah, H. A., Vadivel, B., Shaban, A. A., & Shanmugam, K. (2023). Enhancing English language education: The impact of AI integration in the classroom. *Journal of Humanities and Education Development, 5*(6), 116–123. https://doi.org/10.22161/jhed.5.6.15

Fenilli, L., Ting, S. S. P., Suzuki, A., & Cogo, A. (2025). English as a lingua franca. *The Year's Work in English Studies, 104*(1), 80–95. https://doi.org/10.1093/ywes/maaf105

Fitria, T. N. (2020). Teaching English for specific purposes (ESP) to the students in English language teaching (ELT). *Journal of English Teaching, 5*(1). https://doi.org/10.36456/jet.v5.n01.2020.2276

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research, 25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

García Laborda, J. (2011). Revisiting materials for teaching languages for specific purposes. *3L: The Southeast Asian Journal of English Language Studies, 17*(1), 102–112.

Gass, S. M., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the special issue. *The Modern Language Journal, 82*(3), 299–307.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). Routledge.

Gavrilova, E., & Trostina, K. (2014). Teaching English for professional purposes (EPP) vs content and language integrated learning (CLIL): The case of Plekhanov Russian University of Economics (PRUE). In *Proceedings of the 1st Mediterranean Interdisciplinary Forum on Social Sciences and Humanities (MIFS 2014), Beirut, 23–26 April 2014* (Vol. 2, pp. 7–17).

Ghafar, Z. N., Salh, H. F., Abdulrahim, M. A., Farxha, S. S., Arf, S. F., & Rahim, R. I. (2023). The role of artificial intelligence technology on English language learning: A literature review. *Canadian Journal of Language and Literature Studies, 3*(2), 17–31.

Ghosh, M. (2025). Artificial intelligence (AI) and ethical concerns: A review and research agenda. *Cogent Business & Management, 12*(1), Article 2551809. https://doi.org/10.1080/23311975.2025.2551809

Goh, C. C. M., & Aryadoust, V. (2025). Developing and assessing second language listening and speaking: Does AI make it better? *Annual Review of Applied Linguistics, 45*, 179–199.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105.

González Ramírez, C. (2015). English for specific purposes: Brief history and definitions. *Revista de Lenguas Modernas, 23*, 379–386.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523-547.

Grosman, M. (2008). Slovenščina in drugi jeziki. In M. Ivšek (Ed.), *Languages in education* (pp. 165–175). Zavod Republike Slovenije za šolstvo.

Grosse, C. U. (2004). The competitive advantage of foreign languages and cultural knowledge. *The Modern Language Journal, 88*(3), 351–373.

Grucza, S. (2009). Kategoryzacja języków (specjalistycznych) w świetle antropocentrycznej teorii języków ludzkich [Categorization of languages (for specific purposes) in the light of the anthropocentric theory of human languages]. *Komunikacja Specjalistyczna*, *2*, 15–30.

Gugić, J. (2026). Presentation, ELITE-AI Workshop 'Integrating AI Support in ESP Classes', 7 May 2026. https://bit.ly/4hAm6Y4

Gugić, J., & Blažević, I. (2026). Comparison of AI-supported and traditional English for specific purposes instruction among university students. *Odgojno-obrazovne teme, 9*(1), 1–18. https://doi.org/10.53577/oot.9.1.1

Hashim, H. (2018). Application of technology in the digital era education. *International Journal of Research in Counseling and Education, 1*(2), 1–5.

Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon.

Holliday, A. (2008). Standards of English and politics of inclusion. *Language Teaching, 41*(1), 119–130.

Howatt, A. P. R. (1984). *A history of English language teaching*. Oxford University Press.

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

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Routledge.

Hyland, K. (2014). English for academic purposes. In C. Leung & B. V. Street (Eds.), *The Routledge companion to English studies* (pp. 392–404). Routledge.

Hyland, K., & Jiang, F. K. (2021). Delivering relevance: The emergence of ESP as a discipline. *English for Specific Purposes, 64*, 13–25. https://doi.org/10.1016/j.esp.2021.06.002

Jantakoon, T., Jantakun, T., Jantakun, K., Pongpanich, W., Pasmala, R., Wannapiroon, P., & Nilsook, P. (2025). The effectiveness of artificial intelligence in English instruction for speaking and listening skills: A meta-analysis. *Contemporary Educational Technology, 17*(4), Article ep596. https://doi.org/10.30935/cedtech/17310

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press.

Jenkins, J. (2009). English as a lingua franca: Interpretations and attitudes. *World Englishes, 28*(2), 200–207.

Jenkins, J. (2013). *English as a lingua franca in the international university: The politics of academic English language policy*. Routledge.

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching, 44*(3), 281-315.

Jia, X.-H., & Tu, J.-C. (2024). Towards a new conceptual model of AI-enhanced learning for college students: The roles of artificial intelligence capabilities, general self-efficacy, learning motivation, and critical thinking awareness. *Systems, 12*(3), Article 74. https://doi.org/10.3390/systems12030074

Johns, A. M. (2013). The history of English for specific purposes research. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 5–30). Wiley-Blackwell. https://doi.org/10.1002/9781118339855.ch1

Kachru, B. B. (1996). English as lingua franca. In H. Goebl, P. H. Nelde, Z. Starý, & W. Wölck (Eds.), *Kontaktlinguistik / Contact linguistics / Linguistique de contact: Ein internationales Handbuch zeitgenössischer Forschung* (Vol. 1, pp. 906–913). De Gruyter.

Kankam, D. A. (2025). Inclusive education in a diverse language class: Challenges and best practices. *NECTFL Review, 94*, 117–125.

Khan, S. (2023). *Sal Khan says AI won't destroy education - but there's a catch* [Podcast transcript]. TED. https://www.ted.com/pages/sal-khan-says-ai-wont-destroy-education-transcript

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7–36.

Kim, H. S., Sherman, D. K., & Taylor, S. E. (2008). Culture and social support. *American Psychologist, 63*(6), 518–526.

Klein, W. (1986). *Second language acquisition*. Cambridge University Press.

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

Kompara Lukančič, M. (2023). *English for specific purposes: Selected readings from the field of English for criminal justice and security*. University of Maribor Press.

Kompara Lukančič, M. (2025). Krajšavar: An algorithm for automatic recognition of abbreviations in English dictionary entries compiled in a dictionary of abbreviations. *Lexikos, 35*, 129–142. https://doi.org/10.5788/35-1-2005

Kompara Lukančič, M. (2026a). The automatic recognition of Italian abbreviations: Analysis of Krajšavar and Copilot. *International Journal of Lexicography, 39*, Article ecag006. https://doi.org/10.1093/ijl/ecag006

Kompara Lukančič, M. (2026b). Presentation, ELITE-AI Workshop 'Integrating AI Support in ESP Classes', 7 May 2026. https://bit.ly/4dtdzng

Kompara Lukančič, M., & Smajla, T. (2025). Krajšavar: An algorithm for recognising English abbreviations in texts related to criminal justice and security. *International Journal of Lexicography, 38*(3), 257–269. https://doi.org/10.1093/ijl/ecaf012

KU Leuven. (n.d.). *Responsible use of generative artificial intelligence*. https://www.kuleuven.be/english/education/student/educational-tools/responsible-use-of-generative-artificial-intelligence

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments, 34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Clarendon Press.

Li Destri, E. (2026). Presentation, ELITE-AI Workshop 'Integrating AI Support in ESP Classes', 7 May 2026. https://bit.ly/47FKotE

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence, 9*, Article 100445. https://doi.org/10.1016/j.caeai.2025.100445

Lim, Y. (2024). Classroom heterogeneity and assessment for learning: Evidence from 47 countries using TALIS 2018. *Studies in Educational Evaluation, 83*, Article 101375. https://doi.org/10.1016/j.stueduc.2024.101375

Liu, S., & Alley, F. (2019). Learning from the historical culture of American people for the current society. *Linguistics and Culture Review, 3*(1), 32–47.

Mallette, K. (2014). Lingua franca. In P. Horden & S. Kinoshita (Eds.), *A companion to Mediterranean history*. Wiley-Blackwell. https://doi.org/10.1002/9781118519356.ch21

Martin, F., Gezer, T., & Wang, C. (2019). Educators' perceptions of student digital citizenship practices. *Computers in the Schools, 36*(4), 238–254. https://doi.org/10.1080/07380569.2019.1674621

Mauranen, A., & Ranta, E. (Eds). (2009). *English as a lingua franca: Studies and findings*. Cambridge Scholars Publishing.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

McGilvray, J. (2014). *Chomsky: Language, mind and politics* (2nd ed.). Polity Press.

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO. https://doi.org/10.54675/JKJB9835

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Moeller, A. J., & Catalano, T. (2015). Foreign language teaching and learning. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 327–332). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.92082-8

Mohebbi, A. (2025). Enabling learner independence and self-regulation in language education using AI tools: A systematic review. *Cogent Education, 12*(1), Article 2433814. https://doi.org/10.1080/2331186X.2024.2433814

Morrison, K. (1993). *Planning and accomplishing school-centred evaluation*. Peter Francis.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). *A comprehensive overview of large language models* (arXiv:2307.06435). arXiv. https://doi.org/10.48550/arXiv.2307.06435

Nechifor, A. (2025a). The globitalised discourse of education: Positives and negatives of globalising language acquisition through digitalisation. *Philobiblon: Transylvanian Journal of Multidisciplinary Research in Humanities, 30*(1), 229–247. https://doi.org/10.26424/philobib.2025.30.1.15

Nechifor, A. (2025b). AI-enhanced pragmatic language acquisition in ESP contexts: Connecting grammar development and communicative competence – A comprehensive framework. *Analele Universității "Ovidius" Constanța. Seria Filologie, 36*(2), 1269–1299.

Nechifor, A. (2026a). Presentation, ELITE-AI Workshop 'Integrating AI Support in ESP Classes', 7 May 2026. https://bit.ly/4rTgeg0

Nechifor, A. (2026b). Beyond satisfaction: Examining the integration paradox in AI-enhanced ESP learning. In *INTED2026 Proceedings: 20th International Technology, Education and Development Conference* (Paper 0493). IATED. https://doi.org/10.21125/inted.2026.0493

Nechifor, A., & Dimulescu, C. (2024). Unlocking ESP potential: A culture-infused approach to mobile-assisted language learning. *Analele Universității "Ovidius" Constanța. Seria Filologie, 35*(1), 354–375.

Nechifor, A., Turculeț, A., & Vasilescu, A. (2025). Cross-platform GenAI evaluation for ESP learning scenarios: A computational analysis of linguistic and cultural appropriateness. In *Proceedings of the 20th*

International Conference on Linguistic Resources and Tools for Natural Language Processing (Vol. 20, pp. 162–184). Editura Universității "Alexandru Ioan Cuza" din Iași. https://doi.org/10.47743/ConsILR2025.12

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Ningsih, N., Lenon, S. S., & Rahman, P. (2023). Embracing the wave: The disruption of conversational artificial intelligence toward English for specific purposes (ESP) teaching transformation. *ETERNAL (English, Teaching, Learning, and Research Journal), 9*(2), 260–286.'

Nizzolino, S. (2024). Artificial intelligence in language teaching: Using ChatGPT to assist teachers of English as a foreign language. *Form@re - Open Journal per la formazione in rete, 24*(1), 242–261.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

O'Regan, J. P. (2014). English as a lingua franca: An immanent critique. *Applied Linguistics, 35*(5), 533–552.

Ortega, L. (2014). *Understanding second language acquisition*. Routledge.

Oxford University Press. (n.d.). *Author use of artificial intelligence*. https://academic.oup.com/pages/for-authors/books/author-use-of-artificial-intelligence

Paltridge, B., & Starfield, S. (Eds). (2013). *The handbook of English for specific purposes*. Wiley-Blackwell.

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803–830.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review, 22*, 74–98.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21*(2), 133–148.

Parkinson, J. (2013). English for science and technology. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 155–173). Wiley-Blackwell.

Parra G., L., & Calero S., X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction, 12*(2), 209–226. https://doi.org/10.29333/iji.2019.12214a

Patekar, J. (2017). Common challenges in delivering a successful foreign language lesson. *Strani jezici, 46*(1–2), 101–118.

Pazilah, F. N. P., Hashim, H., & Yunus, M. M. (2019). Using technology in ESL classroom: Highlights and challenges. *Creative Education, 10*(12), 3205–3212. https://doi.org/10.4236/ce.2019.1012244

Peter Lang. (n.d.). *AI statement*. https://www.peterlang.com/ai-statement

Pinter, A. (2019). Agency and technology-mediated task repetition with young learners. *Language Teaching for Young Learners, 1*(2), 139–160. https://doi.org/10.1075/ltyl.00010.pin

Račková, P., & Hošková-Mayerová, Š. (2013). Current approaches to teaching specialised subjects in a foreign language. In *ICERI2013*

Proceedings: 6th International Conference of Education, Research and Innovation (pp. 4775–4783). IATED.

Rahman, M. (2015). English for specific purposes (ESP): A holistic review. *Universal Journal of Educational Research, 3*(1), 24–31.

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology, 37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). (2016). *Official Journal of the European Union, L 119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union, L 2024/1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). (2026). *Official Journal of the European Union, L 2026/1744*. https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng

Repka, R. (2020). Towards a definition of language. *Philologia, 30*(1), 149–170.

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning*. Cambridge University Press.

Riordan, E. (2018). The medium and the object: Foreign language classroom discourse. In E. Riordan, *Language for teaching purposes*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-71005-1_4

Robinson, P. (1991). *ESP today: A practitioner's guide*. Prentice Hall.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed.). Blackwell.

Sahito, J. K. M., Panwar, A. H., & Ramzan, I. (2025). Exploring the impact of artificial intelligence (AI) on the listening skills of English as a second language (ESL) learners. *Journal of Applied Linguistics and TESOL, 8*(1), 1059–1067.

Sahlender, M., Helbert, S., ten Hagen, I., Knaus, A., & Weiss, Z. (2024). Investigating spoken classroom interactions in linguistically heterogeneous learning groups: An interdisciplinary approach to process video-based data in second language acquisition classrooms. *Applied Corpus Linguistics, 4*(3), Article 100104. https://doi.org/10.1016/j.acorp.2024.100104

Samarin, W. J. (1987). Lingua franca. In U. Ammon, N. Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society* (Vol. 1, pp. 371–374). Walter de Gruyter.

Saussure, F. de. (1966) [1976]. *Course in general linguistics* (W. Baskin, Trans.). McGraw-Hill.

Saville-Troike, M., & Barto, K. (2017). *Introducing second language acquisition* (3rd ed.). Cambridge University Press.

Schmitz, J. R. (2017). English as a lingua franca: Applied linguistics, Marxism, and post-Marxist theory. *Revista Brasileira de Linguística Aplicada, 17*(2), 335–354. https://doi.org/10.1590/1984-6398201710866

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 133–158.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Shah, S. H. R., Altaf, A. R., & Mughal, K. H. (2025). Exploring the impact of AI tools on reading proficiency among undergraduate students. *The Critical Review of Social Sciences Studies, 3*(3), 663–670.

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*. Advance online publication.

Shiukashvili, T. (2020). Comparative-contrastive analysis of anthropomorphic figures of evil in English and Georgian languages. *International Journal of Linguistics, Literature and Culture, 6*(1), 24–31.

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74–87. https://doi.org/10.4018/JDM.2020040105

Silitonga, L. M. (2024). Boosting students' ESP vocabulary by utilising AI chatbot. *Eternal: English Teaching Journal, 15*(2), 275–283. https://doi.org/10.26877/eternal.v15i2.605

Smajla, T., & Kompara Lukančič, M. (2025). Slovenian tourism students' views on the use of artificial intelligence in LSP learning. *Tourism in Southern and Eastern Europe, 8*, 89–104. https://doi.org/10.20867/tosee.08.9

Smajla, T., & Podovšovnik, E. (2026). Students' views on the use of artificial intelligence in language for specific purposes (LSP) courses. *International Journal of Instruction, 19*(2), 125–146. https://doi.org/10.29333/iji.2026.1927a

Son, J.-B., Ružić, N. K., & Philpott, A. (2025). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning, 5*(1), 94–112. https://doi.org/10.1515/jccall-2023-0015

Spada, N. (2019). Classroom observation research. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 186–207). Cambridge University Press.

Stanford University, Office of Community Standards. (n.d.). *Generative AI policy guidance*. https://communitystandards.stanford.edu/generative-ai-policy-guidance

Starfield, S. (2013). Historical development of language for specific purposes. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0505

Stockwell, G. (2007). A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL, 19*(2), 105–120.

Strevens, P. (1980). *Teaching English as an international language: From practice to principle*. Pergamon Press.

Sudipa, I. N., Aryati, K. F., Susanta, I. P. A. E., & Anggayana, I. W. A. (2020). The development of syllabus and lesson plan based on English for occupational purposes. *International Journal of Psychosocial Rehabilitation, 24*(4), 290–300.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Synekop, O., Lytovchenko, I., Lavrysh, Y., & Lukianenko, V. (2024). Use of Chat GPT in English for engineering classes: Are students' and teachers' views on its opportunities and challenges similar? *International Journal of Interactive Mobile Technologies, 18*(3).

Tafazoli, D., Gómez María, E., & Huertas Abril, C. A. (2019). Intelligent language tutoring system: Integrating intelligent computer-assisted language learning into language education. *International Journal of Information and Communication Technology Education, 15*(3). https://doi.org/10.4018/IJICTE.2019070105

Teodorescu, A. (2010). Teaching English for specific purposes. *Petroleum-Gas University of Ploiești Bulletin, Philology Series, 62*(2), 67–74.

Tessensohn, T. C., Yunus, M. M., & Ismail, H. H. (2025). Using AI-powered tools in enhancing reading skills in the ESL classroom: A systematic review (2020–2024). *International Journal of Academic Research in Progressive Education and Development, 14*(2), 57–70.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*(2), 91–112. https://doi.org/10.1093/applin/4.2.91

Tsao, C.-H. (2011). English for specific purposes in the EFL context: A survey of student and faculty perceptions. *Asian ESP Journal, 7*(2), 126-149.

Tudor, I. (1997). LSP or language education? In R. Howard & G. Brown (Eds.), *Teacher education for languages for specific purposes* (pp. 90-102). Multilingual Matters.

Uber Grosse, C., & Voght, G. M. (2012). The continuing evolution of languages for specific purposes. *The Modern Language Journal, 96*(s1), 190–202. https://doi.org/10.1111/j.1540-4781.2012.01304.x

University of Cambridge, Blended Learning Service. (n.d.). *AI policy framework*. https://blendedlearning.cam.ac.uk/artificial-intelligence-and-education/ai-policy-framework

University of Oxford. (n.d.). *Generative AI at Oxford*. https://www.ox.ac.uk/gen-ai

Vadivel, B., Shaban, A. A., Ahmed, Z. A., & Saravanan, B. (2023). Unlocking English proficiency: Assessing the influence of AI-powered language learning apps on young learners' language acquisition. *International Journal of English Language, Education and Literature Studies, 2*(6), 55–62. https://doi.org/10.22161/ijeel.2.6.7

VanPatten, B. (2003). *From input to output: A teacher's guide to second language acquisition*. McGraw-Hill.

VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition* (2nd ed.). Bloomsbury.

VanPatten, B., & Williams, J. (Eds.). (2015). *Theories in second language acquisition: An introduction* (2nd ed.). Routledge.

Vincent, J., Md Yunus, M., & Said, N. E. M. (2025). Using AI platforms to improve listening and speaking skills in ESL primary students. *International Journal of Academic Research in Progressive Education and Development, 14*(1), 1829–1843. https://doi.org/10.6007/IJARPED/v14-i1/24848

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.

Widodo, H. P. (2016). Teaching English for specific purposes (ESP): English for vocational purposes (EVP). In W. A. Renandya & H. P. Widodo (Eds.), *English language teaching today: Linking theory and practice* (pp. 277–291). Springer. https://doi.org/10.1007/978-3-319-38834-2_19

Willis, J. (1996). *A framework for task-based learning*. Longman.

Woods, A. (2018). American culture: A sociological perspective. *Linguistics and Culture Review, 2*(1), 1–12.

Wragg, E. C. (1994). *An introduction to classroom observation*. Routledge.

Wu, X., & Li, R. (2026). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *The Asia-Pacific Education Researcher, 35*(4), 703–714. https://doi.org/10.1007/s40299-024-00853-2

Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126*, Article 103498.

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomised controlled trial. *Humanities and Social Sciences Communications, 12*, Article 425. https://doi.org/10.1057/s41599-025-04672-8

Xie, Q. (2016). Developing communicative competence for the globalised workplace in English for occupational purposes course in China. *Journal of Language Teaching and Research, 7*(6), 1142–1152.

Xu, H., Teo, H. H., & Tan, B. C. Y. (2006). Information privacy in the digital era: An exploratory research framework. In *Proceedings of the Twelfth Americas Conference on Information Systems* (Paper 120, pp. 900–906). Association for Information Systems. https://aisel.aisnet.org/amcis2006/120/

Yang, H. (1986). A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing, 1*(2), 93–103.

Ypsilandis, G. S., & Kantaridou, Z. (2007). English for academic purposes: Case studies in Europe. *Revista de Lingüística y Lenguas Aplicadas, 2*.

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). *A survey of large language models* (arXiv:2303.18223). arXiv. https://doi.org/10.48550/arXiv.2303.18223

Zhu, M., & Wang, C. (2025). A systematic review of artificial intelligence in language education: Current status and future implications. *Language Learning & Technology, 29*(1), 1–29. https://hdl.handle.net/10125/73606

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–12). Routledge.

Annex 1

Teacher Observation Sheet

* Indicates required question

1. University: *

Mark only one oval.

- UNIUD
- UNIUM
- UNIPU
- UNITBV

2. Teacher: *

Mark only one oval.

- Costantini, Francesco (UNIUD)
- Gallitelli, Eleonora (UNIUD)
- Li Destri, Emanuela (UNIUD)
- Kompara Lukančič, Mojca (UNIUM)
- Smajla, Tilen (UNIUM)
- Gugić, Jelena (UNIPU)
- Dimulescu, Cristina (UNITBV)
- Nechifor, Andreea (UNITBV)

3. Date: *

Example: January 7, 2019

4. Number of observation sheet per partner:

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20

5. Number of students in the group: *

6. Duration of the class in minutes: *

Mark only one oval.

- 45'
- 90'
- 100'

7. Study level: *

Mark only one oval.

- BA 1st Year
- BA 2nd Year
- BA 3rd Year
- MA 1st Year
- MA 2nd Year

8. Field: *

Mark only one oval.

- tourism
- mobility
- law
- criminal justice
- education
- social media marketing
- sport

9. Lesson stage: *

Check all that apply.

- [ ] Lead-in
- [ ] Teaching
- [ ] Practice

10. Topic: *

_____

11. Objectives - Warming-up / Lead-in stage:

Check all that apply.

- [ ] To activate learners' interest in reading specialised texts
- [ ] To activate learners' interest in writing for dedicated specialised activities
- [ ] To activate learners' interest in specialised vocabulary aquisition
- [ ] To activate learners' interest in grammar knowledge in context
- [ ] To activate learners' interest in listening to specialised audio materials (video included)
- [ ] To activate learners' interest in speaking in appropriate, particular specialised contexts

12. Objectives - Teaching:

Check all that apply.

- [ ] To teach learners' how to read specialised texts and extract meaningful inormation from them
- [ ] To teach learners' how to writing dedicated specialised texts
- [ ] To teach learners' how to use specialised vocabulary in dedicated, specialised contexts
- [ ] To teach learners' how to activate the most appropriate grammar strucutres in context
- [ ] To teach learners' how to listen to specialised audio materials (video included) and extract meaningful inormation from them
- [ ] To teach learners' how to speak in professional settings

13. Objectives - Practicing:

Check all that apply.

- [ ] To develop reading skills
- [ ] To develop writing skills
- [ ] To develop vocabulary dedicated to professional settings
- [ ] To develop listening skills
- [ ] To develop speaking fluency
- [ ] To develop grammatical competence for dedicated professional settings

14. Language acquisition tool: *

15. Materials used alongside the AI tools: *

Check all that apply.

- [ ] Laptop / Desktop Computer
- [ ] Mobile Phones / Tablets
- [ ] Internet Connection (Wi-Fi / LAN)
- [ ] Interactive Whiteboard
- [ ] Traditional Whiteboard / Markers
- [ ] Video Projector
- [ ] Headphones / Earbuds
- [ ] Microphones
- [ ] Speakers / Sound System
- [ ] VR Headset (e.g., Oculus, HTC Vive)
- [ ] Printed Worksheets / Handouts
- [ ] USB Drive / External Storage
- [ ] Document Camera / Visualizer
- [ ] Paper and Pens
- [ ] Classroom Management Software (e.g., Google Classroom)
- [ ] External Monitor / TV Screen
- [ ] Flipchart
- [ ] Recording Device (Camera / Voice Recorder)
- [ ] Smartboard
- [ ] Printer / Scanner

16. Level of language knowledge: *

Mark only one oval.

- ( ) A1 – Beginner
- ( ) A2 – Elementary
- ( ) B1 – Intermediate
- ( ) B2 – Upper Intermediate
- ( ) C1 – Advanced
- ( ) C2 – Proficient
- ( ) Mixed-ability group
- ( ) Unknown / Not assessed yet

Use of AI tools

17. Purpose of AI tool use - LISTENING

Check all that apply.

- [ ] Podcast/audio content generation
- [ ] Speech-to-text transcription
- [ ] Audio speed adjustment
- [ ] Accent/dialect variation generation
- [ ] Listening comprehension question creation
- [ ] Dictation exercise generation
- [ ] Pronunciation model audio
- [ ] Authentic listening material curation
- [ ] Audio subtitle generation
- [ ] Other

18. Purpose of AI tool use - SPEAKING

Check all that apply.

- [ ] Conversation simulation/practice
- [ ] Speech recognition and feedback
- [ ] Pronunciation assessment
- [ ] Fluency practice
- [ ] Role-play scenario generation
- [ ] Oral presentation outline
- [ ] Interview preparation dialogue
- [ ] Accent reduction feedback
- [ ] Real-time speaking correction
- [ ] Other

19. Purpose of AI tool use - READING

Check all that apply.

- [ ] Text generation for reading practice
- [ ] Reading comprehension question design
- [ ] Text simplification/level adjustment
- [ ] Authentic material adaptation
- [ ] Content summarization
- [ ] Text annotation and glossing
- [ ] Reading speed tracking
- [ ] Genre-specific text creation
- [ ] Professional document templates
- [ ] Other

20. Purpose of AI tool use - WRITING

Check all that apply.

- [ ] Writing prompt generation
- [ ] Essay outline creation
- [ ] Grammar and spelling correction
- [ ] Style and register feedback
- [ ] Paraphrasing assistance
- [ ] Citation formatting
- [ ] Blog/article creation
- [ ] Email and correspondence templates
- [ ] Academic writing feedback
- [ ] Plagiarism detection
- [ ] Other

21. Purpose of AI tool use - GRAMMAR

Check all that apply.

- [ ] Grammar explanation generation
- [ ] Error correction and analysis
- [ ] Contextualized grammar examples
- [ ] Grammar exercise creation
- [ ] Syntax pattern identification
- [ ] Tense and aspect practice
- [ ] Sentence structure analysis
- [ ] Grammar rule simplification
- [ ] Other

22. Purpose of AI tool use - VOCABULARY

Check all that apply.

- [ ] Specialized/technical vocabulary lists
- [ ] Contextual word usage examples
- [ ] Collocation identification
- [ ] Synonym and antonym generation
- [ ] Terminology database creation
- [ ] Vocabulary quiz generation
- [ ] Word family and derivation
- [ ] Register-appropriate vocabulary
- [ ] Flashcard creation
- [ ] Etymology and word origin explanation
- [ ] Other

23. Use of AI tools: *

Mark only one oval.

- ( ) under-use
- ( ) appropriate
- ( ) over-reliance

24. Learning: *

Mark only one oval.

- ( ) diminished
- ( ) maintained
- ( ) enhanced

Student Engagement

25. Active Participation *

Mark only one oval per row.

	often	sometimes	never
Initiating interaction	( )	( )	( )
Volunteering	( )	( )	( )
Applying themselves to the tasks	( )	( )	( )
Seeking clarifications	( )	( )	( )
Solving tasks	( )	( )	( )

26. Frequency of interaction with peers *

Mark only one oval.

- ( ) high
- ( ) medium
- ( ) low

27. Degree of involvement *

Mark only one oval.

- active by volition
- active by request
- passive

28. Type of interaction *

Mark only one oval per row.

	Often	Sometimes	Never
Teacher–Student			
Student–Student (peer interaction)			
Student–AI Tool / Platform			
Whole-class participation			
Group work / Collaboration			
One-on-one (e.g., during tutoring)			

29. Use of Target Language *

Mark only one oval per row.

	diminished	maintained	increased
Frequency			

30. Motivation *

Mark only one oval.

- High
- Developing
- Low

Student-AI Interaction

31. Independence in using AI *

Mark only one oval.

- Student proactively seeks help from AI
- Student uses AI as directed by teacher
- Student avoids using AI

32. Quality of interaction *

Mark only one oval per row.

	yes	sometimes	no
Communication with AI			
Use of prompts			
Interpretation of AI feedback			

33. Critical Evaluation of AI Output *

Mark only one oval per row.

	yes	sometimes	no
Students verify AI accuracy because they are aware of hallucinations			
Students verify AI accuracy because they suspect wrong / inappropriate / incomplete / false answers			
Studens accept AI output without questioning			
Not the case, as there were instances when we did not work with chatbots, so this check was not needed			

34. Integration of AI into Learning Process *

Mark only one oval per row.

	yes	sometimes	no
Students use AI in isolation			
Students struggle to connect AI use to learning			
Students rely on AI support too much			

35. **Language Skills** *

Mark only one oval per row.

	improved	developing	diminished
Listening comprehension			
Speaking skills			
Reading comprehension			
Writing skills			

36. Language Skills *

Mark only one oval per row.

	improved	developing	diminished
Grammar			
Specialised vocabulary			

37. **Class behavior and management** *

Mark only one oval per row.

	improved	developing	diminished
Asking for help			
Interaction with teacher			
Adherence to instructions			
Interaction with the tools			
Avoiding distractions			
Respect for others			

Google Forms

Annex 2

Student Self-Assessment Sheet: In-Class Activities

* Indicates required question

1. University: *

Mark only one oval.

- University of Udine (UNIUD) – Italy
- University of Maribor (UNIUM) – Slovenia
- Juraj Dobrila University of Pula (UNIPU) – Croatia
- Transilvania University of Brașov (UNITBV) – Romania

2. Instructor: *

Mark only one oval.

- Costantini, Francesco (UNIUD)
- Gallitelli, Eleonora (UNIUD)
- Li Destri, Emanuela (UNIUD)
- Kompara Lukančič, Mojca (UNIUM)
- Smajla, Tilen (UNIUM)
- Gugić, Jelena (UNIPU)
- Dimulescu, Cristina (UNITBV)
- Nechifor, Andreea (UNITBV)

3. Date: *

_____

Example: January 7, 2019

4. Number of observation sheet per partner (PLEASE ASK TEACHER):

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30

5. Study level: *

Mark only one oval.

- BA 1st Year
- BA 2nd Year
- BA 3rd Year
- MA 1st Year
- MA 2nd Year

6. Field: *

Mark only one oval.

- tourism
- mobility
- law
- criminal justice
- education
- social media marketing
- sport

7. Effectiveness of AI tool use (How did AI help or not help me?) *

Mark only one oval per row.

	AI makes no difference in my learning style	AI makes it harder for me to learn	AI makes it easier for me to learn
Learning			

8. Effectiveness of AI tool use (How did AI help or not help me?) *

Mark only one oval per row.

	I used it too little	I used it appropriately	I used it too much
Use of AI tools			

9. Active Participation (How much did I participate?) *

Mark only one oval per row.

	often	sometimes	never
Speaking up without being called on			
Starting conversations			
Responding to questions			
Asking for my teacher's help when I didn't understand			
Asking for AI's help when I didn't understand			

10. Interaction with peers (How did I work with classmates?) *

Mark only one oval per row.

	high	medium	low
Frequency of interaction with my colleagues			

11. Interaction with peers (How did I work with classmates?) *

Mark only one oval per row.

	Worked well together	Interacted just as much as usual	Didn't interact much
Quality of interaction			

12. Use of Target Language (How much did I use the language we are learning?) *

Mark only one oval per row.

	very frequently	sometimes	never
How frequently did I use the language we are learning?			

13. Use of Target Language (How much did I use the language we are learning?) *

Mark only one oval per row.

	yes	so and so	not really
Was I able to get my message across?			
Did I use correct grammar?			

14. Motivation (How did I feel about learning today?) *

Mark only one oval.

- I feel I did well and was motivated
- I feel I was trying and improving
- I feel I struggled with motivation

15. Independence in using AI (How did I decide to use AI?) *

Mark only one oval.

- I proactively sought help from AI
- I used AI because the teacher told me to
- I avoided using AI

16. Quality of interaction (How well did I communicate with the AI and understand it?) *

Mark only one oval per row.

	effective	partially effective	ineffective	no chatbot used
How effective were my interactions with AI?				
How effective were my instructions or questions to the AI?				
How effectively did I understand and use what the AI told me?				

17. Critical Evaluation of AI Output (Did I check if the AI was right?) *

Mark only one oval.

- I verified AI output because I am familiar with the concept of hallucinations
- I verified AI output because I suspected something was not correct
- I accepted AI output without questioning
- Not the case, as there were instances when we did not work with chatbots, so this check was not needed

18. Integration of AI into Learning Process (How did AI fit into my learning?) *

Mark only one oval.

- AI felt like a useful tool for in class-learning
- I used AI for a specific task but it didn't connect much to my overall in class-learning
- I struggled to connect AI use to in-class learning; I didn't understand how AI helped me learn

19. Overall Feeling *

Mark only one oval.

- I found AI integration helpful
- I found AI integration useless
- I found AI integration detrimental

Google Forms

Annex 3
Student Self-Assessment Sheet: At-Home Activities

* Indicates required question

1. Email *

2. Date: *

Example: January 7, 2019

3. University: *

Mark only one oval.

- University of Udine (UNIUD) – Italy
- University of Maribor (UNIUM) – Slovenia
- Juraj Dobrila University of Pula (UNIPU) – Croatia
- Transilvania University of Brașov (UNITBV) – Romania

4. Instructor: *

Mark only one oval.

- Costantini, Francesco (UNIUD)
- Gallitelli, Eleonora (UNIUD)
- Li Destri, Emanuela (UNIUD)
- Kompara Lukančič, Mojca (UNIUM)
- Smajla, Tilen (UNIUM)
- Gugić, Jelena (UNIPU)
- Dimulescu, Cristina (UNITBV)
- Nechifor, Andreea (UNITBV)

5. Study level: *

Mark only one oval.

- BA 1st Year
- BA 2nd Year
- BA 3rd Year
- MA 1st Year
- MA 2nd Year

6. Field: *

Mark only one oval.

- tourism
- mobility
- law
- criminal justice
- education
- social media marketing
- sport

7. Name the AI tool used during the semester for at home language activities, homework assigned for the language classes, personal educational purposes (Check all that apply):

Check all that apply.

- [ ] Claude
- [ ] Gemini
- [ ] NotebookLM
- [ ] Google AI Studio
- [ ] Perplexity
- [ ] ChatGPT
- [ ] Copilot
- [ ] Grammarly
- [ ] Classtime
- [ ] ZenoChat
- [ ] DeepSeek
- [ ] CoralAI
- [ ] Other (please specify in the next question)

8. If you used "Other" in the previous question, please specify:

9. What did you use the AI tool for (Check all that apply): *

Check all that apply.

- [ ] Brainstorming for ideas
- [ ] Rephrasing paragraphs
- [ ] Checking grammar
- [ ] Improving style
- [ ] Conversation
- [ ] Power point slides generation
- [ ] Infographics / Mindmaps
- [ ] Language scenarios in dedicated contexts
- [ ] Flash card generation for checking knowledge based on dedicated input
- [ ] Grammar drilling in context
- [ ] Specialised vocabulary building
- [ ] Quizzes
- [ ] Translation
- [ ] Collocation identification and learning
- [ ] Proofreading
- [ ] Checking for accuracy
- [ ] Summarising
- [ ] Content comparing
- [ ] Term extraction
- [ ] Hallucination observation
- [ ] Other (please specify in the next question)

10. If you used "Other" in the previous question, please specify:

11. Effectiveness of AI tool use (How did AI help or not help me?) *

Mark only one oval per row.

	AI made no difference in my learning style	AI made it harder for me to learn	AI made it easier for me to learn
Learning:			

12. Effectiveness of AI tool use (How did AI help or not help me?) *

Mark only one oval per row.

	I used it too little	I used it appropriately	I used it too much
Use of AI tools:			

13. Use of AI tools for language self-improvement during the semester *

Mark only one oval per row.

	frequently	sometimes	never
How much did I use AI tools for my homework?			
How much did I use AI tools for English language self-improvement?			
How much did I use AI tools for specialised vocabulary self-improvement?			

14. Motivation (How did I feel about learning during the semester?) *

Mark only one oval.

- I feel I did well and was motivated
- I feel I was trying and improving
- I feel I struggled with motivation

15. Independence in using AI (How did I decide to use AI?) *

Check all that apply.

- [ ] I proactively sought help from AI
- [ ] I used AI because the teacher told me to
- [ ] I avoided using AI

16. Quality of interaction (How well did I communicate with the AI and understand it?)*

Mark only one oval per row.

	effective	partially effective	ineffective	no chatbot used
How effective were my interactions with AI?	◯	◯	◯	◯
How effective were my instructions or questions to the AI?	◯	◯	◯	◯
How effectively did I understand and use what the AI told me?	◯	◯	◯	◯

17. Critical Evaluation of AI Output (Did I check if the AI was right?)*

Mark only one oval.

- ◯ I verified AI output because I am familiar with the concept of hallucinations
- ◯ I verified AI output because I suspected something was not correct
- ◯ I accepted AI output without questioning
- ◯ Not the case

18. Integration of AI into the Learning Process (How did AI fit into my learning?)*

Check all that apply.

- [ ] I incorporated AI into study habits
- [ ] AI felt like a useful tool for at-home self-learning
- [ ] I used AI for a specific task but it didn't connect much to my overall at-home-learning
- [ ] I struggled to connect AI use to self-learning; I didn't understand how AI helped me learn

19. I think that I integrated AI generated content in my assignments ethically*

Mark only one oval.

- ◯ yes
- ◯ so and so
- ◯ not really

20. Overall feeling *

Mark only one oval.

- ◯ I found AI integration helpful
- ◯ I found AI integration useless
- ◯ I found AI integration detrimental

21. I feel that the activities assisted by AI tools made me progress in acquiring specialized vocabulary * *Mark only one oval.*

	1	2	
True	◯	◯	False

22. I feel that the activities assisted by AI tools made me understand the use of grammar structures in specific contexts

Mark only one oval.

	1	2	
True	◯	◯	False

23. I feel that the activities assisted by AI tools made me relate to pragmatic concepts in real-life situations / conversations

Mark only one oval.

	1	2	
True	◯	◯	False

24. I feel that the activities assisted by AI tools made me integrate professional writing practices in my learning process

Mark only one oval.

	1	2	
True	◯	◯	False

25. I feel that the activities assisted by AI tools made me develop my reading / listening and speaking skills

Mark only one oval.

	1	2	
True	◯	◯	False

26. My participation in the experimental group made me feel

Mark only one oval.

- ◯ Frustrated for fear I might be behind with the information provided to the control group in a traditional manner
- ◯ Special, because I was part of an innovative learning context
- ◯ Neutral with respect to either feelings of superiority or inferiority during the language acquisition process

Google Forms