New Directions in Foreign Language Education

Pedagogies for a Changing World

Book Series

Contemporary Studies on Digital Humanities, Vol. III

Ali Tilbe, Yusuf Topaloğlu, Gökhan Dinar (Eds.)

Yunus Emre Sarı & Ali Tilbe (Eds.)

New Directions in Foreign Language Education

Pedagogies for a Changing World

maurer.press
Frankfurt am Main

Bibliographic Information of the German National Library

The German National Library lists this publication in the German National Bibliography; detailed bibliographic data are available online at the German National Library website: http://dnb.d-nb.de

© Published by Verlag Hans Jürgen Maurer

under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

The work, including all of its parts, is protected by copyright. Any use beyond the narrow limits of copyright law without the consent of the copyright holder is prohibited and punishable by law. This applies in particular to reproductions, translations, and processing in electronic systems.

The book has undergone a double blind peer review.

Published in cooperation with AYG Press İstanbul and IUC University Press

maurer.press

is an imprint of
Verlag Hans-Jürgen Maurer
Im Trierischen Hof 14
60311 Frankfurt/Germany
www.maurer.press
GPSR-Kontakt: info@maurer.press

ISBN 978-3-949550-54-6
e-ISBN 978-3-949550-55-3

TABLE OF CONTENTS

PART I: PEDAGOGICAL FRAMEWORKS

Teachers as Developers: AI-Assisted Coding in Language Education

Yunus Emre Sarı *15*

Reconceptualizing EFL Pedagogy in the Age of Generative AI and Digitalization

Aysel Deregözü, İlayda Erdem *41*

Savoring Language Through Translanguaging: A Conceptual and Discursive Perspective

Maghsoud Hassani Mehr *67*

PART II: THE TEACHER

Learning to See Before Learning to Teach: The Unnamed Role in the Practicum of Pre-Service Teachers of German in Türkiye

Ebru Üstün, Bilal Üstün *89*

Motivation to Learn with GenAI: A Self-Determination Theory Perspective on Pre-Service German Language Teachers

Hamdullah Şahin *111*

AI as Support, Not Substitute: English Instructors' Perceptions of ChatGPT in Foreign Language Teaching

Şule Çınar Yağcı, Tuğba Aydın Yıldız *133*

The Transformation of the Teacher's Role in the Age of Artificial Intelligence: A Didactic Decision Tree Proposal Based on a Prompt Engineering Taxonomy

Hakan Dolan *165*

PART III: MATERIALS AND TASK DESIGN

A Stage-Oriented Model of AI-Personalized German Language Preparation from A1 to B2

Betim Neyran Can *189*

AI-Driven Learning Analytics for French Language Materials Design across CEFR levels A1 to B1

Fayrouz Merabti *211*

Artificial Intelligence-Based Dynamic Persona in Language Education: An Examination Of Sociocultural Context And Pragmatic Competence

Burcu Ün, Gizem Arslan *231*

From AI Translation to Pragmatic Post-Editing: A Task-Based Framework for Developing University EFL Learners' Pragmatic Competence

Maryam DelfarianTurk *255*

PART IV: ASSESSMENT AND ETHICS

Assessment in French Language Instruction in the Age of Generative Artificial Intelligence

Gizem Köşker *281*

Living Portfolio in Iterative Artificial Intelligence Feedback and the Reorientation of Foreign Language Assessment

Alireza Navidmoghaddam *305*

Ethical Dimensions of Artificial Intelligence in Foreign Language Education

Zeyneb Dolan *331*

Preface

We both work in foreign language education: Yunus Emre Sarı in German language teaching and Ali Tilbe in French language teaching. The book began with a simple question we put to colleagues who do the same work: where did they think our field was going?

We did not start from artificial intelligence or from any other single theme. We chose the subtitle, *Pedagogies for a Changing World*, before a single chapter arrived, because we wanted the contributors to show us what that changing world looked like.

When the chapters came in, artificial intelligence was everywhere. We had not asked for it, but we were not surprised either. It had already become part of how colleagues talk about materials, assessment, teacher education, feedback and classroom practice.

The book also keeps in view questions that no technology has settled: what learners feel while they learn, and how future teachers learn to see a classroom before they teach in one. For that reason the subtitle still seemed right to us. Change in foreign language education is wider than any one technology.

The fourteen chapters were written by eighteen colleagues at eight universities across Türkiye. Ten of them teach English, French, German and Italian in schools of foreign languages; eight prepare future teachers of English, French and German in faculties of education. They do not give one answer to where the field is going, and we did not expect them to. What they show is what a group of teachers and teacher educators considers worth examining now.

We think of the book as a conversation among colleagues. One colleague names a problem, and another suggests a way of looking at it.

We hope it will be useful to language teachers, teacher educators, researchers and everyone else working in foreign language education. It is published open access so that the conversation can reach beyond the institutions and classrooms where it began.

We thank all our authors for their work and their patience, and Verlag H.J. Maurer in Frankfurt am Main for publishing the book and making it freely available.

Yunus Emre Sarı & Ali Tilbe

Introduction

Language teachers have always had to decide what their learners should be able to do and how to tell whether they can. What changes from one period to the next is the setting in which those decisions are made. Today the most visible change is a technology that drafts and redrafts text, answers questions about grammar, produces code, reads learner data and keeps a conversation going for as long as anyone likes. It does not settle a single pedagogical question. Teachers still have to judge when such help is appropriate, what learners should manage on their own, how a generated sentence should be checked and which parts of the work cannot be handed over.

New Directions in Foreign Language Education: Pedagogies for a Changing World starts from these judgements and from the older questions behind them. Its chapters range over post-method pedagogy, positive affect, the practicum, prompt writing, personalised materials, assessment and ethics. What they share is a concern with how foreign language education should respond when the conditions it was built on change.

We have arranged the fourteen chapters in four parts: pedagogical frameworks, the teacher, materials and task design, and assessment and ethics. As we see it, this is roughly the order in which these questions come up in practice.

Part I. Pedagogical Frameworks

The book opens with Yunus Emre Sarı's "Teachers as Developers: AI-Assisted Coding in Language Education" (Chapter 1), which looks at the gap between using a digital material and building one. Sarı separates two kinds of work: building a working artefact, and judging whether it is pedagogically and technically sound. For some small artefacts, AI-assisted coding makes building easier, but the standards for judging the result stay the same. From this difference the chapter develops a typology of teacher-built artefacts and a five-item checklist that runs from writing the prompt to a short note after first use. It also asks what happens once an artefact exists: maintenance and abandonment, accessibility, learner data, ownership, and artefacts that work but teach little.

Aysel Deregözü and İlayda Erdem (Chapter 2) ask what remains of post-method pedagogy once the learning environment itself adapts to the learner. Their chapter, "Reconceptualizing EFL Pedagogy in the Age of Generative AI and Digitalization", argues that generative AI might put post-method principles into practice, among them learner autonomy, context-sensitive exposure to language and wider learning opportunities. The authors also take the costs seriously and raise questions of data bias, digital equity and pedagogical integrity. In their account, AI works as adaptive scaffolding within a context that teachers and learners shape together.

Maghsoud Hassani Mehr (Chapter 3) turns to a process the field has paid little attention to: savoring, the way learners attend to a positive experience, intensify it and make it last. "Savoring Language Through Translanguaging: A Conceptual and Discursive Perspective" brings together research on positive psychology, second language affect and translanguaging. Hassani Mehr argues that translanguaging provides the semiotic architecture through which savoring is enacted, and he builds a model of "savoring-through-translanguaging" around a four-stage Discursive-Savoring Cycle, with a coding matrix for analysing it. The chapter distinguishes savoring from enjoyment and treats it as discursive work that learners do with their whole linguistic repertoire.

Part II. The Teacher

This part follows the teacher's career from its start. Ebru Üstün and Bilal Üstün (Chapter 4) examine the practicum hours that the regulation requires to be planned but does not name. "Learning to See Before Learning to Teach: The Unnamed Role in the Practicum of Pre-Service Teachers of German in Türkiye" is about the time pre-service teachers spend in classrooms without teaching. Drawing on the 2021 national regulation, four institutional practicum documents and research on professional vision, the authors show that being present in a classroom does not by itself lead to professional learning. It does so only where observation is structured, and what student teachers notice depends on what they already know. Their framework organises observation along the object of perception, the form of mediation and the observer's epistemic stance, and moves from imitative to analytical and then inquiry-oriented observation.

Hamdullah Şahin (Chapter 5) stays with student teachers of German and asks what motivates them to learn with generative AI. "Motivation to

Learn with GenAI: A Self-Determination Theory Perspective on Pre-Service German Language Teachers" reports survey data from 94 pre-service teachers. They show high autonomous motivation and low levels of both controlled motivation and amotivation. The strong negative correlation between intrinsic motivation and amotivation suggests that engagement with these tools is associated more closely with interest than with external pressure. Şahin concludes that teacher training needs a hybrid model that addresses intrinsic and extrinsic motivational needs together.

The next chapter moves from student teachers to practising ones. Şule Çınar Yağcı and Tuğba Aydın Yıldız (Chapter 6) interviewed four English instructors after a four-hour professional development programme and report their views in "AI as Support, Not Substitute: English Instructors' Perceptions of ChatGPT in Foreign Language Teaching". The instructors valued AI for developing materials, finding resources and supporting learner autonomy. They were just as quick to name their concerns about privacy, academic integrity, student dependency and reduced cognitive engagement. In their view, responsible use still depends on teacher expertise and clear institutional guidance.

Hakan Dolan (Chapter 7) closes the part. He treats prompt engineering as a didactic competence. In "The Transformation of the Teacher's Role in the Age of Artificial Intelligence: A Didactic Decision Tree Proposal Based on a Prompt Engineering Taxonomy" he reinterprets a 41-technique taxonomy for the needs of language teaching and builds a scenario-based decision tree that helps teachers choose among six techniques for source-grounded material development, from document-grounded prompting and retrieval-augmented generation to chain-of-verification. One point runs through the chapter: checking by the model does not replace the teacher's review, which must always follow it.

Part III. Materials and Task Design

Betim Neyran Can (Chapter 8) questions the common assumption that more personalisation means more automation. "A Stage-Oriented Model of AI-Personalized German Language Preparation from A1 to B2" links the type and intensity of AI support to proficiency, task purpose, learner independence and teacher judgement. At A1 and A2, support stays controlled and close to the curriculum. At B1 and B2, it can extend to comparison, revision, mediation, register awareness and reflective feedback.

Personalisation, in this model, is therefore calibrated rather than simply increased.

Fayrouz Merabti (Chapter 9) asks what learner data can actually contribute to materials design. In "AI-Driven Learning Analytics for French Language Materials Design across CEFR levels A1 to B1" she proposes a cycle in which learner interaction produces data, AI helps analyse it and teachers use the results to revise their materials, with a different emphasis at A1, A2 and B1. Analytics should inform design decisions, she argues, and should not make them. Alignment with the CEFR, cultural relevance and pedagogical appropriateness remain the teacher's responsibility.

Burcu Ün and Gizem Arslan (Chapter 10) start from a familiar weakness of textbook dialogues: they leave out social and emotional variation. "Artificial Intelligence-Based Dynamic Persona in Language Education: An Examination Of Sociocultural Context And Pragmatic Competence" uses characters generated by large language models to bring that variation into everyday settings such as restaurants, family, health and work. Learners then have to adapt their language to interlocutors who differ in social position, personality and emotional state. The authors describe AI in this role as a "sociocultural laboratory" for exploring pragmatic and sociolinguistic variation.

Maryam DelfarianTurk (Chapter 11) also works on pragmatic competence, but through revision. In "From AI Translation to Pragmatic Post-Editing: A Task-Based Framework for Developing University EFL Learners' Pragmatic Competence" she treats AI-generated and machine-translated language as a draft to be evaluated and rewritten before it is accepted. Her five-dimensional framework and six-stage Pragmatic Post-Editing Cycle direct learners' attention beyond grammatical fluency to communicative purpose, the relationship with the recipient, directness and imposition, politeness and the writer's own voice.

Part IV. Assessment and Ethics

Gizem Köşker (Chapter 12) addresses the assessment problem directly. When learners can generate, translate and revise texts with AI, an accurate final text does not necessarily show what they can do on their own. In "Assessment in French Language Instruction in the Age of Generative Artificial Intelligence" she proposes a four-stage process: independent

initial performance, documented AI-supported editing, critical validation, and an oral explanation in which the student transfers the work to a new context. The chapter includes sample tasks and analytical scoring criteria for A2, B1 and B2.

Alireza Navidmoghaddam (Chapter 13) reads seven empirical studies, which are usually discussed separately, as one body of evidence. They cover portfolio writing and assessment, biometric-informed reading, a longitudinal speaking corpus, task-based instruction and learner emotion. "Living Portfolio in Iterative Artificial Intelligence Feedback and the Re-orientation of Foreign Language Assessment" proposes that sustained, proficiency-calibrated feedback within an iterative, documented cycle may support linguistic, cognitive and affective development. The author presents this as a proposition that still has to be tested and states its limits openly: pragmatic competence, the governance of longitudinal data and a narrow evidence base.

Zeyneb Dolan (Chapter 14) closes the book. "Ethical Dimensions of Artificial Intelligence in Foreign Language Education" discusses privacy and data security, bias, academic integrity, risks to human interaction, digital inequality and the opacity of algorithmic decisions. Drawing on the governance frameworks and ethical guidelines of UNESCO, the European Commission and the World Economic Forum, Dolan argues for principles-based governance combined with digital literacy training, so that AI supports teachers instead of replacing them.

Taken together, the chapters do not propose a single model for the future of foreign language education. They place their emphasis differently: on teacher competence and professional vision, on motivation and affect, on materials, on pragmatic interaction, on assessment or on governance. Yet one concern keeps returning, and that is human judgement. Whatever the tool or task, decisions about purpose, appropriateness, responsibility and educational value remain with teachers, learners and institutions.

Part I
Pedagogical Frameworks

Teachers as Developers: AI-Assisted Coding in Language Education

Yunus Emre Sarı [1]

Abstract
Creating interactive digital materials has often required either working within the constraints of authoring tools or drawing on programming expertise. AI-assisted coding begins to change this, since teachers can now obtain working artefacts by describing them in ordinary language. Easier production leaves two questions open: whether an artefact is worth building and whether it serves its learners. The chapter therefore separates two kinds of work: building an artefact and judging its pedagogical and technical adequacy, both when it is planned and after it is built. For some small artefacts, AI-assisted coding can reduce the expertise and effort needed to build them. It does not lower the standards by which the result must be judged, even where tools perform some checks, and a working artefact can hide a failure to meet them. The chapter proposes a typology of teacher-built artefacts organised by whether each receives learner input and what it does with it, and examines the costs that arise once a working version exists: maintenance and abandonment, accessibility, learner data, ownership, and artefacts that work but teach little. It closes with a five-item checklist covering planning, testing, and first use, and with implications for teacher education.
Keywords: foreign language education, AI-assisted coding, materials development, teacher-made materials, computer-assisted language learning

Introduction

Materials development is an established field within language teaching, with its own literature and evaluative criteria (Tomlinson, 2012). Teachers have long taken part in this work by selecting, adapting, and rewriting materials. When the material is software, however, teachers have often worked within the options provided by authoring tools. Moving beyond those options has typically required programming knowledge or expertise in more complex authoring environments, which cannot be assumed to form part of every language teacher's preparation.

The need for teacher-made materials varies across languages and teaching contexts. Even in the saturated English language teaching market,

[1] Assoc. Prof. Dr., Istanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education, Department of Foreign Language Education, German Language Teaching Program, ORCID: 0000-0002-3589-8055, *y.sari@iuc.edu.tr*, Türkiye

commercial pressures constrain what a global coursebook can contain (Mishan, 2022). Teachers of less commonly taught languages often face a more basic problem, since materials suited to their teaching context may be unavailable or out of date (Özçelik & Kennedy Kent, 2023). In such contexts, producing materials can be a practical necessity.

Generative artificial intelligence (AI) extends what teachers can produce. For example, Shin et al. (2026) describe no-code routes through which English language teachers can build and revise their own applications, and Mørch et al. (2025) report two case studies of teachers adapting and configuring language models. In a chapter published elsewhere, the author examined the use of generative AI in skill-based German instruction (Sarı, 2025). This conceptual chapter turns to teachers who build digital artefacts with AI assistance rather than adjust an existing tool or model. It develops an argument from selected literature and an illustrative exercise; it does not report a systematic review or a study of learning outcomes. It examines teacher-built software in terms of pedagogical justification and learner fit, drawing on materials development and computer-assisted language learning (CALL) scholarship (Chapelle, 2001; Tomlinson, 2012), and considers the responsibilities that arise once a working artefact exists.

The chapter's central point is that building an artefact and judging it are two different kinds of work. Building requires knowledge, skills, time, and resources. Judging requires the expertise and effort to decide whether the artefact is pedagogically and technically adequate, both while it is being planned and once a working version exists. For some small artefacts, AI-assisted coding can reduce what building requires. It does not lower the standards for judging the result, and whether it reduces the effort of judging remains open. For teachers without programming expertise, ideas that exceeded the limits of available templates could remain unbuilt. Pedagogical design questions still mattered, but there was no working version to evaluate. When a working version is easy to obtain, successful production can be mistaken for successful design. A functioning artefact does not, by itself, show that it suits the intended proficiency level, serves a worthwhile learning objective, is accessible to its learners, or handles their data appropriately. Nor can teachers assume that a coding assistant will raise these questions. Where requirements remain unstated, the assistant

may either settle consequential design decisions without drawing attention to them or leave them open.

The chapter first places teacher-built software within materials development, CALL authoring, and end-user development, and explains the difference between building and judging. It then considers what AI-assisted coding has changed and what it has left in place, proposes a typology of teacher-built artefacts illustrated by an exercise built with an assistant, and examines the costs that follow. It closes with implications for teacher education, among them a five-item checklist that runs from writing the prompt to a short note after first use.

Theoretical Background

Materials Development and Teachers as Designers

Materials development rests on a normative commitment: a material needs a defensible relationship to how learners learn and to the purposes it serves (Tomlinson, 2012). Principled decision-making therefore matters in selecting and adapting existing materials as well as in creating new ones. Tomlinson's (2023) account of blended learning connects materials design to meaningful language use, affective and cognitive engagement, and opportunities for communication. These principles provide grounds for evaluating what a digital artefact asks learners to do, beyond whether its interface works. Recent research on generative AI-supported materials development also highlights the continuing importance of teachers' contextual judgement (Lo, 2025).

Kali et al. (2015) frame teachers as designers of technology-enhanced learning and treat teacher design as professional knowledge work in its own right. McKenney et al. (2015) situate teacher design knowledge within an ecological framework of design assets and needs, noting that teacher education rarely develops design expertise beyond lesson planning. Teaching involves design reasoning, whose objects are goals, sequences, learner activity, and the artefacts that carry them. A further question is whether the design knowledge and resources developed through creating lessons, tasks, and texts transfer unchanged to software design.

Authoring Systems and End-User Development

CALL has a long history of authoring tools that allow teachers to create materials without programming. Drawing on Otto and Pusack's (2009) account of these tools, this chapter distinguishes building a new artefact from adjusting one that someone else has designed. Their account treats authoring as a continuum. Simple templates, predefined forms into which content is typed, sit at one end. Complex authoring environments sit at the other end and offer more flexibility, but they demand advanced skill, considerable time, and resources. The same authors identify several barriers to teacher authoring: demands on time and resources, training requirements, a lack of professional recognition, and the short lifespan of the resulting software. At the template end, convenience comes from what has been defined in advance, but even a skilled teacher cannot go beyond it. The available interaction types are those the designer of the template anticipated, and a need outside that set cannot be met by filling in the form differently. With such templates, teachers could already create materials without programming, but their options were bounded by each template's design. Production through conversation loosens that bound.

When a teacher builds an artefact, what it does and the order of its steps come from the teacher's description, although an assistant may fill in details the teacher has not specified. When a teacher adjusts an existing tool, the teacher changes settings and prompts within a structure designed by someone else. Some cases fall between the two. Mørch et al. (2025) illustrate why this distinction needs care: the teachers in their two cases mainly adapted and adjusted existing tools, although one project also involved building a system together with researchers and computer scientists. The distinction used here concerns the teacher's role in specifying what the artefact does; building can still make use of existing tools and services.

End-user development refers to the methods and tools that let people who are not professional software developers create or modify a software artefact (Lieberman et al., 2006). Ko et al. (2011) open their survey of end-user software engineering with the observation that most programs are written not by professional developers but by people whose expertise lies in other domains and who need computational support for their own purposes. That is the position this chapter assigns to the teacher. Ko et al. add that end-user developers face many of the same software engineering

challenges as professionals, including understanding requirements, making design decisions, testing, and debugging. For two decades, end-user software engineering has documented the obligations that come with building software with spreadsheets and web-authoring tools. In materials development, those obligations rest on different ground. Quality depends not only on whether the software works correctly but also on pedagogical justification. The teacher responsible for developing the material is accountable to learners: an artefact can run exactly as specified yet serve no defensible learning objective.

Building and Judging

These strands provide complementary accounts of teachers' design work. Materials development addresses the selection, adaptation, creation, and evaluation of learning resources; teacher-design research examines learning environments; CALL authoring extends these concerns to digital tools; and end-user development explicitly includes software created by non-professionals. The question here is how the evaluative responsibilities these strands describe apply when a teacher can obtain a working digital artefact by describing to a coding assistant what it should do.

AI-assisted coding makes it useful to separate two kinds of work that are easily treated as a single technical requirement. Creating a working artefact requires knowledge, skills, time, and resources; until recently, anything beyond a template also required programming or comparable technical skill. For some small, self-contained artefacts, AI assistance reduces these demands, because the assistant supplies code for behaviour that a template does not offer. Judging the result still requires the expertise and effort to decide whether it is pedagogically and technically adequate, both in the decisions made beforehand and in the finished artefact. Easier creation does not lower the standards for that judgement, and tools that help with checking may change this work without removing it. Responsibility for the judgement remains with people. That judgement is pedagogical, covering the level an artefact addresses, the objective it serves, and the learner performance it expects. It is also technical, covering accessibility, security, and data handling.

Sarkar and Drosos (2025) analysed videos in which programmers built software through conversation with an AI assistant. Programming expertise remained necessary, but the work shifted towards managing what the

assistant was given, checking generated code quickly, and deciding when to intervene directly. However, none of the sessions featured a user without programming expertise, and the authors do not extend the finding to such users. Carrying it over to teachers is the author's inference. On that inference, a teacher may need additional support when a judgement depends on aspects of the artefact that cannot be established by trying the page.

What AI-Assisted Coding Has Changed

The most immediate change is how easily a teacher can reach a first working version. A bespoke artefact used to sit at the far end of the authoring continuum. Consider an exercise tool that records which form in a case paradigm a learner got wrong, rather than merely recording that a response was incorrect, and presents that form again in fresh lexical contexts until the learner gets it right consistently. A template that records only correctness cannot provide this diagnosis; a template can do so only if its design supports the relevant error categories and sequencing. Moving beyond such a template requires a more flexible authoring environment or code.

Two controlled comparisons in general software engineering point in opposite directions, and both are preprints, not refereed reports. Peng et al. (2023), some of whose authors work for the tool's vendor, studied an assistant that suggested code as programmers typed, not the conversational loop at issue here, and found that professional programmers completed a small programming task, begun from a partly prepared project, markedly faster. Becker et al. (2025) found the opposite with experienced developers working on large projects they already knew well: with the tools, their tasks took longer, yet afterwards the developers still believed the tools had made them faster. Neither study measured the production of teaching materials, so their relevance here is analogical.

For teachers without programming expertise, the change is that some bespoke artefacts may now be within reach. This concerns access to production, not simply the speed of an existing workflow. It remains open whether teachers who work by prompting and trying the page misjudge the assistant's effect on their time, and whether they also underestimate the faults in what it produces or the need to check it.

A second change is the speed of trial and revision. Teacher design has been theorised as iterative and situated (Kali et al., 2015; McKenney et al., 2015). When a small revision can be made through a conversational turn, a teacher can test an intuition against a working artefact. Technical difficulty also screens out fewer projects than it did. It once kept some projects from reaching learners, but that barrier reflected effort rather than pedagogical value. With that filter weakened, more artefacts can reach learners, and problems that nobody looked for beforehand surface in use. An artefact that does not run is usually noticed quickly, whereas one that runs without error but does no defensible teaching work is noticed only by someone who is already asking the pedagogical question.

Teachers' existing design knowledge does not guarantee that they will catch these problems. A digital artefact fixes decisions about which learner responses are treated differently, what stays unchanged, and when a person reviews its output. A teacher can adjust a lesson during interaction; an artefact can adjust only in the ways specified in advance. A working artefact can also exist before its rationale, which makes it easy to justify the finished result after the fact.

The shift that Sarkar and Drosos (2025) describe also leaves much of the work in place, and some of that work requires expertise beyond pedagogical design. Specifying requirements is still a human act. When a teacher leaves part of the description unspecified, the assistant may fill the gap in a way that does not serve the intended pedagogy. Whether the artefact actually fits what learners need remains a materials-development judgement, and the assistant cannot make it for the teacher. Maintenance remains an obligation, and CALL's own sustainability literature separates one-off artefacts from projects that survive technological and institutional change (Kennedy & Levy, 2009).

An informal observation from the author's own teaching illustrates the concern. Pre-service teachers were asked to introduce digital applications and then teach a short lesson with one. Most chose ready-made platforms, and a few built something of their own with a coding assistant. The work repeatedly stalled before the teaching step, and where an artefact was built, the level, the age group, and the objective still had to be settled after the artefact existed. The difficulty also arose with ready-made platforms, suggesting that building software alone did not explain it.

Producing a useful teaching artefact draws on pedagogical experience, opportunities to practise design, familiarity with the tools, and access to technical support. AI-assisted coding can reduce some technical demands for small artefacts. However, it does not supply all of these conditions. Mørch et al. (2025, Introduction section, para. 2) observe that "not all teachers are interested, able, or willing to utilize the increased agency." How much checking an artefact needs depends largely on what it does with learner input.

A Typology of Teacher-Built Artefacts

The proposed typology distinguishes four functional categories according to whether an artefact receives learner input and what it does with that input. These categories are not stages on a scale of complexity, and one artefact can combine several functions, as the exercise below does. The first judges closed responses; the second evaluates open production; the third sustains an exchange without assigning scores; and the fourth transforms teacher-supplied material without receiving learner responses. Gamification is treated here as an optional extra, most often added to closed-response practice, not as a separate function.

The scope is set separately and covers artefacts whose output reaches learners, either directly or through materials and feedback the teacher passes on. Tools used solely for teacher or researcher analysis fall outside it, as do environments whose purpose is communication among people. All four categories can be implemented at different points on the authoring continuum, and AI-assisted coding has brought its far end within reach of some teachers who do not program, at least for small artefacts. The possibility of moving beyond templates means that the teacher as a customiser of existing CALL materials (Barcomb & Cardoso, 2025) no longer marks the upper limit of what teachers can build.

Practice and Feedback Tools

The first and most familiar category presents items and judges responses; it may also give different follow-up items depending on the kind of error made. Its value lies in matching practice more closely to a particular group of learners. A teacher may face a cohort whose error profile is not addressed by available materials, and where the market is thin, suitable

material may be absent or dated (Özçelik & Kennedy Kent, 2023). One decision that shapes the quality of such a tool is the level label, which is a claim about the proficiency an item targets. In a study of alignment to the Common European Framework of Reference for Languages (CEFR), expert judges assigning levels to test items varied in the criteria and descriptors they used and in their interpretations of the levels (Harsch & Hartig, 2015).

That study concerns testing, and carrying it over to teaching materials is the author's analogy, resting on a shared task: in both cases a person judges an item's level before any learner has answered it. A level label produced with a generative model is therefore a design claim in need of justification, and an item does not acquire its level by being generated. A practice tool can also respond only to the kinds of error its designer anticipated, and a generative model may produce plausible items whose level labels are unsupported, without reliably signalling the mismatch. Figures 1 to 4 show how a simpler tool, one that explains each error but gives every learner the same items, was specified and built.

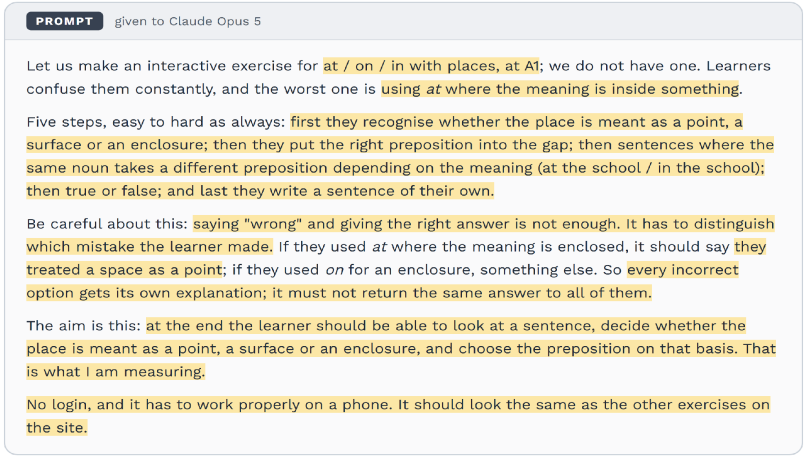


Figure 1. The prompt given to Claude Opus 5: a teaching specification for an interactive exercise on *at*, *on*, and *in* with places. It is reproduced in full; it states the teaching requirements and asks that the exercise look like the other exercises on the site. Highlighting marks each decision the teacher stated.

The prompt is a specification written in the language of teaching rather than of programming. It names the structure (*at*, *on*, and *in* with places),

the CEFR level (A1), and the confusion that matters most. It sets out five steps from recognition to a sentence of the learner's own, states what the learner should be able to do at the end, and asks that every incorrect option receive its own explanation of the mistake made. Its analysis of a place as a point, a surface, or an enclosure is close to cognitive-linguistic accounts of prepositions, on which computer-based tutors for English prepositions have already been built (Wong et al., 2018; Zhao et al., 2020). Figures 2 to 4 show the exercise that the assistant produced from the prompt.

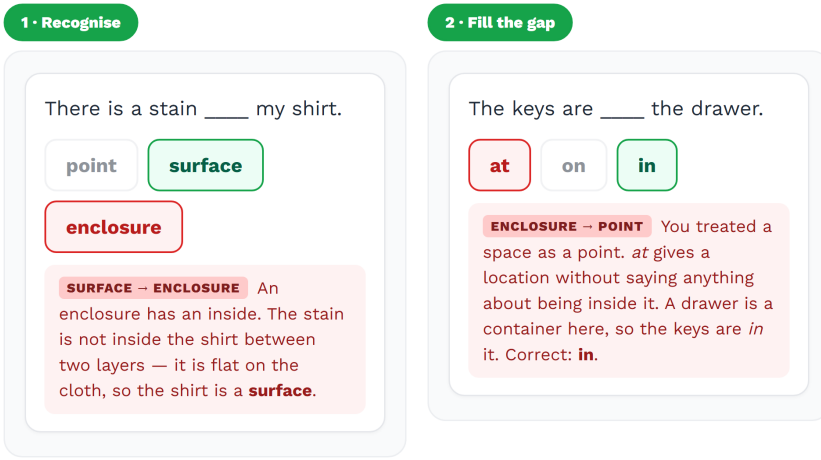


Figure 2. Steps 1 and 2 of the exercise produced from the prompt in Figure 1, each after an incorrect answer: deciding how the place is meant, then choosing the preposition. Figures 2 to 4 show the version the assistant first returned, before corrections were requested; the item numbers shown on the page within each step are omitted.

The first two steps separate meaning from form. The sequence asks the learner to decide whether a place is a point, a surface, or an enclosure before choosing the word, and a wrong preposition is then treated as a misreading of the place. The assistant carried this into the feedback, as the prompt required. Each wrong option has its own explanation, headed by a label that pairs the sense of the place with the sense implied by the answer, which is more than the correctness feedback used as a comparison condition in one such tutor (Zhao et al., 2020). Even so, the items need a teacher's reading. An item from step 2 not shown here reads: *They are ____ the airport now, waiting for their flight*. The answer *in* is marked

wrong although the explanation itself calls it close. Figure 3 shows the next two steps.

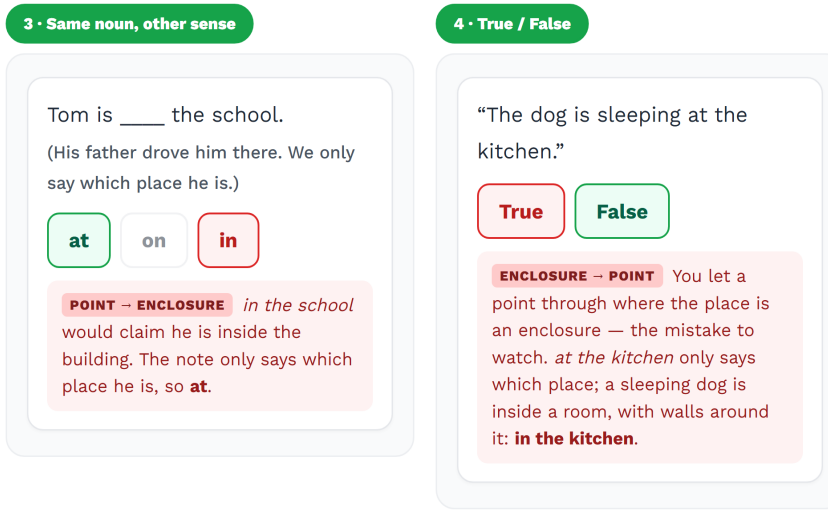


Figure 3. Steps 3 and 4, each shown after an incorrect answer: first, a noun whose sense is given in brackets and which returns in the next item with another sense; second, a judgement on a complete sentence. First version, as returned by the assistant.

Steps 3 and 4 move from choosing to contrasting and judging. Because the same noun takes two prepositions in neighbouring items, the learner cannot simply remember which word goes with which noun, and step 4 asks for the same decision without offering the prepositions to choose from. The explanation for the dog in the kitchen identifies the confusion that the prompt called the worst one. The bracketed cues also need a teacher's review. The cue in step 3, *We only say which place he is*, is incomplete as a locative expression, so an A1 learner meets a faulty model sentence in the cue and again in the explanation. The item shown in Figure 3 carries a further problem. With *school*, English marks the institutional reading by dropping the article, as in *at school*. Once the article is there, both *at the school* and *in the school* remain available, and only the bracketed cue rules the second one out, while the feedback labels the answer a confusion of point and enclosure. This contrast came from the teacher's own specification (Figure 1), so the problem is already present in the prompt. Figure 4 shows the fifth and last step.

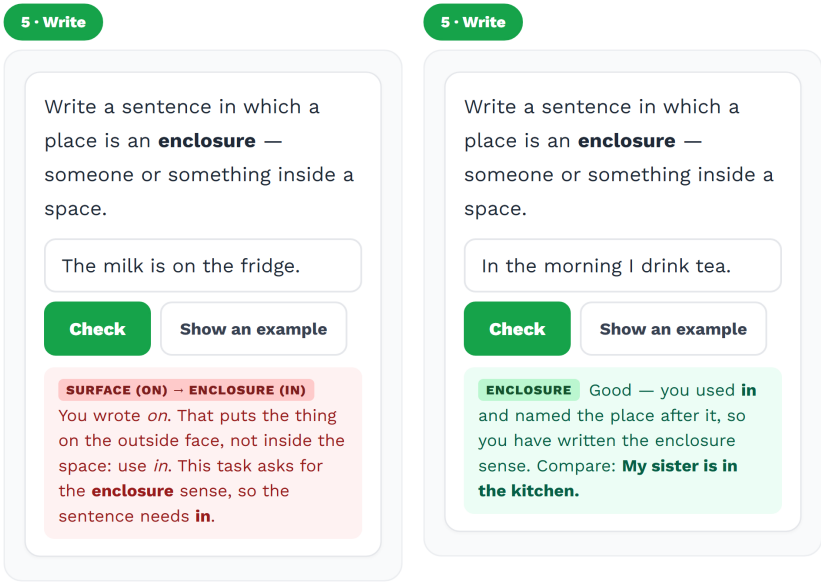


Figure 4. Step 5, the writing task that asks for a place meant as an enclosure. The two sentences were entered by the author to test the check. The first receives corrective feedback; the second, which uses *in* for a time of day, is accepted. First version, as returned by the assistant.

Step 5, where learners write a sentence of their own, goes furthest beyond choosing, but its check does not follow the aim of the prompt. The second sentence in Figure 4, *In the morning I drink tea*, names a time rather than a place, yet the page accepts it with the feedback "Good — you used *in* and named the place after it, so you have written the enclosure sense." In this test, the page fails to distinguish the required spatial use of *in* from a non-spatial one, although that distinction is what the prompt set out to measure. The judgement therefore misses the intended construct, that is, what the task is meant to capture (Knoch, 2011). The flaw becomes evident only when someone enters a sentence that fails to express the required spatial meaning. The distinctions of the first four steps came from the teacher and were included in the feedback, even though an item such as the airport sentence applies them in a context that leaves the sense open. The fifth step keeps their labels and reverses the order of the two senses that the arrow pairs, yet the sentence it accepts in Figure 4 does not express the distinction those labels name. In judging a sentence the learner has written, step 5 already does the work of the second category described

below, but without establishing whether the sentence expresses the required meaning. The prompt never said how such a sentence should be judged, so this step needed human review before its feedback could be treated as an assessment of meaning.

The level label is another matter. The specification states A1 without giving a reason; the explanations use English that an A1 learner may not follow, partly because the specification itself asked for the labels point, surface, and enclosure, and nothing in the production of the exercise checked whether its items belong at that level.

A tool that judges responses can also be made to reward them, and gamified skill tools take that step by adding points or other rewards. They attach a game mechanic to a narrow grammatical or lexical regularity, such as timed sorting for gender assignment, and can be designed to support repeated practice. Because such a tool still judges a closed response, it belongs with the practice tools. The decision this adds is which mechanic fits which regularity and what the scoring rewards.

Engagement and motivation recur as the rationale for gamification in second language learning, yet the review assembling that rationale found no study linking particular elements to the outcomes claimed (Dehghanzadeh et al., 2021). Outside language learning, a meta-analysis found that gamification has positive but small effects, largest on what learners learn and smallest on their behaviour (Sailer & Homner, 2020). The main risk is drift: a mechanic that rewards speed makes rapid recognition sufficient for a high score, even where the objective is accurate production.

Assessment and Feedback Tools

The second category reads open learner production, in writing or speech, to support the teacher's review. It can reduce the work of providing feedback, but what it should do depends on the assessment criteria. Knoch (2011) argues that the decisions taken in developing a rating scale, above all the choice of criteria, are decisions about what the assessment defines as its construct, and that theory does not settle them on the developer's behalf.

Knoch's argument predates generative AI and concerns human raters, so extending it to a teacher-built rubric is an argument by analogy. A rubric that one teacher builds and applies alone may borrow descriptors from

established scales, but it does not automatically include moderation of borderline cases or checks on agreement between human and automated judgements. In practice, one person has defined the construct, and it will be reconsidered only if someone deliberately reviews it later. Unless outputs are reviewed against the intended criteria, what the assessment measures in practice may narrow to whatever the tool actually scores.

Open-Ended Interaction Tools

The third category takes learner input as turns in an exchange, as in a role play, a simulated dialogue, or a conversation partner, and does not score those turns as responses. It can offer more chances for sustained interaction where class size and contact hours limit them. Two decisions are central to the pedagogical quality of such a tool. The first is what counts as a successful turn and what the partner does with errors: whether it repairs an error, lets it pass, or flags it for the teacher. Even without scoring, this defines the kind of interaction the tool is intended to support. Interactional competence has both cognitive and social sides, and turning it into something that can be scored is difficult even for human raters (Galaczi & Taylor, 2018).

The second decision concerns the instructional sequence that structures the exchange. Dong et al. (2026) describe custom chatbots for teaching genre analysis, whose prompts and workflows were refined across four instructional phases under continuous human oversight. The study examined design and use rather than learning effectiveness. That case falls between building and adjusting: the instructional sequence originated with its designers, but it was realised by adjusting existing chatbots. Fluent turns can continue even when the exchange has drifted from its instructional purpose. A teacher-built partner cannot be relied on to signal such drift unless its builder has designed that signal in.

Materials Adaptation Tools

The fourth category takes no learner input at all. It converts a single source into versions at several levels or in several formats, which saves work for a teacher with several proficiency levels in the room. Producing content with generative AI is a different matter from building such a tool: prompting a model to simplify one text produces content. Making a tool that accepts any text a colleague submits, takes it through steps the teacher has

fixed, and returns levelled versions is building a tool; adjusting an existing one lies between the two.

The main design decision is what must stay the same in every version: the communicative function, the target structures, or the task demand. Stating it is a judgement that materials development places at its centre. A second decision is which version reaches which learner. The level-labelling problem of the first category applies here as well. A single prompted output passes under a teacher's eye, but 30 versions produced automatically do not unless someone deliberately reviews them.

What the Four Categories Share

None of the four categories requires the teacher to write the code unaided, and in each case judging the result is both a pedagogical and a technical task. Each category demands a specification of the level, the objective, and the point of human inspection. Where the artefact takes learner input, it also needs the expected learner product; where it transforms a source, it needs a statement of what must stay the same. What the artefact returns must then be evaluated. The typology therefore describes the decisions teachers still have to make, and AI assistance does not make all of them easier.

Judging is not equally demanding across the four categories. Closed items are easier to check, which may lead teachers to favour closed-response practice. Where acceptable answers can be specified in advance, scoring can be checked against a reviewed answer key. This may make that part of evaluation more manageable, but it does not establish an item's pedagogical value or remove the need for accessibility, security, and data-handling checks. Open production and sustained interaction require additional judgements about meaning and the abilities the activity is intended to elicit.

There is also a language-related difference between artefacts that call on a model each time they are used and those that use it only beforehand. Some artefacts pass learner input, source material, or a request for new items to a model each time they are used, so that every output depends on how the model performs in the target language. In others, the model is used only beforehand, to make the page and perhaps its items before any

learner arrives. What it produced is then fixed, and the items it wrote can be read once before use.

Assessment and feedback tools, interaction tools, and adaptation tools may rely on a model each time they are used, although they can also work by fixed rules. Judging open meaning is harder than checking a closed answer, as the fifth step of the exercise shows (Figure 4). A practice tool needs a model only beforehand unless it creates items or feedback while the learner works; otherwise, any errors of level sit in a fixed set of items that a teacher can read, however fluent those items appear. Where the second, third, and fourth categories do rely on a model during use, their output depends on how that model handles the target language. In a language that models handle poorly, what such an artefact returns needs closer checking, and a thin materials market does not by itself show how well a model performs in a language.

The Costs of Teacher-Built Artefacts

Published material can come with piloting, editing, level validation, legal clearance, accessibility compliance, version management, and a publisher answerable for the result. When the teacher becomes the developer and no institution shares the load, that work falls largely on the builder. The resulting costs are maintenance and abandonment, accessibility, learner data, ownership, and artefacts that work but teach little. The last is the one that the checklist proposed below is built to catch by trying the page; the others can stay out of sight while the page works.

The first cost is maintenance. Of the obstacles to teacher authoring that Otto and Pusack (2009) named before generative models, the short lifespan of software remains. The evidence on abandonment comes from other domains and is only an analogy. In 133 widely used open-source projects, about two-thirds depended on one or two people who knew how the software worked, and a third on a single person (Avelino et al., 2016); a teacher who builds alone may be the only person responsible for maintaining the artefact.

In CALL, the long-running programme that Kennedy and Levy (2009) analyse was the work of a team with institutional support. A teacher-built page rarely has such a team. When it stops working, any report of the

problem reaches the teacher, if it reaches anyone, and a page kept on an outside website depends on arrangements the teacher does not control.

A page can also stop working for reasons outside it. Some pages ask an AI model for a response each time a learner uses them, for instance to comment on a sentence the learner has written. Such a page depends on a service the teacher does not control: if the provider changes the model or its terms, or withdraws the service, the page behaves differently or stops, although nobody has touched it.

Maintenance becomes a lasting cost when an artefact remains in use after the course or by people other than its original teacher, holds learner data, depends on an outside AI service, or cannot be rebuilt without the original prompt. Where none of these conditions holds, building something disposable is defensible.

Accessibility brings a different cost. Learners who use screen readers, work with the keyboard alone, or need strong contrast are in language classrooms too, and pages made by an assistant do not reliably provide for them, especially when the prompt does not ask for accessibility. When web pages were generated with and without a request for accessibility, expert reviewers found far more problems in those made without it, whereas an automated checker found only a small difference, in the opposite direction (Guriță & Vatavu, 2025). Asking for accessibility in the prompt therefore helps, but a clean automated report does not show that a page is accessible. Checking a page for learners who cannot see it or cannot use a mouse is specialist work beyond a language teacher's training.

Learner data raise a separate problem. A page that accepts learner writing, records speech, or logs errors handles data that may stay on the device, go to an institutional server, or be passed to an outside service. Studies have repeatedly found security weaknesses in what AI assistants produce: in one in four to three in ten samples taken from real projects (Fu et al., 2025) and, in a preprint not yet peer reviewed, in nine of every ten publicly available applications built through conversation (Deng et al., 2026). Neither study concerns teacher-built materials. Such weaknesses do not show when a teacher tries the page. The rules that govern educational data also differ by country and institution, and a trial run cannot settle which of them apply.

Ownership raises questions at two levels. The first is the artefact as an output of employment. In US higher education, institutions may treat instructional materials as work made for hire, whereas academic tradition treats the creator as the owner (Kranch, 2008). The second is the generated code. Assistants sometimes reproduce code that others have published under a licence, often enough for this to be a documented risk (Yu et al., 2023), and under US copyright law, according to the Copyright Office's reading of current technology, prompts alone do not make a person the author of what is generated (U.S. Copyright Office, 2025). Whether a teacher may license a page for colleagues therefore depends on who owns it and on the terms attached to what it reuses. Wiley and Hilton (2018) set out the permissions to retain, reuse, revise, remix, and redistribute a resource, and they build a pedagogy on those permissions. If permissions carry pedagogical weight in this way, choosing a licence is also a pedagogical decision, since the choice determines whether colleagues may adapt the artefact.

The final cost is pedagogical. An artefact can look convincing because its interface responds normally even when the activity lacks a defensible relationship to its learning objective. A teacher who builds through conversation sees the page, not how it works, and the fact that it works can draw more attention than what it teaches. A pedagogical rationale can be examined before use. Claims about learning effects also require evidence from learners. Materials development already provides a way to address this: materials are justified against principles, not by their appearance (Tomlinson, 2012), and teacher production is a reflective act (Bouckaert, 2019). Evaluation, however, is now easier to postpone until after the first working version.

A written record of the teacher's decisions also helps after use: set against what the artefact actually did, it serves as a short check. Learners have a stake in this evaluation as well. Once a teacher-built artefact contributes to a grade, transparency and appeal become live questions. Whether learners are told that the tool was made by their teacher, and whether the judgement it produces stays open to review by a person, are design decisions that a coding assistant cannot be relied on to raise.

Implications for Teacher Education

A programming course might seem the obvious curricular response, but it mainly addresses building, which AI-assisted coding has already made easier for some small artefacts. Evaluating whether a task is appropriate for its stated proficiency level and supports its learning objective requires pedagogical judgement. Such judgement is central to both the teacher-as-designer role (Kali et al., 2015; McKenney et al., 2015) and materials development (Tomlinson, 2012). Determining whether an artefact sends learner data elsewhere or can be used with a screen reader is technical work that design preparation does not by itself supply. Teacher education should develop pedagogical judgement and basic technical literacy, while technical verification may require specialist support beyond ordinary teacher preparation.

Building through conversation can begin before the pedagogical and technical decisions discussed above have been made. Fixing the objective and the expected outcome before selecting means is a long-standing precept of instructional planning and of materials development (Tomlinson, 2012), and CALL has its own criteria for judging a task, among them language learning potential, learner fit, and practicality (Chapelle, 2001). In the workflow considered here, the teacher applies these criteria without reading code: the teacher writes a prompt, tries the page that comes back, and asks in the same conversation for whatever is wrong to be changed. What the prompt leaves unsaid, the assistant may decide without asking, and the teacher sees such a decision only if trying the page brings it to light.

The five items below form a proposed checklist that follows this way of working. The first two belong in the prompt, items 3 and 4 come before learners use the artefact, and item 5 continues after first use. If one of items 1 to 3 cannot yet be settled, the artefact waits. Table 1, which follows the list, sets the items against the exercise in Figures 1 to 4.

1. **Level and aim in the prompt.** The prompt states the level, what learners should be able to do afterwards that they could not do before, and why available material does not already serve that aim. For a tool that rewrites texts, it also states what must stay the same in every version.

2. **How answers are judged.** The prompt states what counts as a right answer and what the learner is told after a wrong one. Where learners write or speak freely, the teacher decides whether the artefact judges the answer at all or a person reviews it. Anything that counts towards a grade is reviewed by a person, and learners are told so.
3. **Trying it with answers that should fail.** Before learners see it, the teacher works through the artefact as a learner would, entering correct answers phrased in other ways, wrong answers, and answers that make no sense, and reads what the artefact says each time. Faults are described in the same conversation, and the artefact is tried again after each change.
4. **A second look, where possible.** Preferably, a colleague who teaches the level also works through the artefact before learners do.
5. **Keeping the prompt.** The teacher keeps the prompt, the corrections asked for, and, after first use, a short dated note on what learners did with the artefact. Together these form a written rationale; without them, neither the teacher nor a colleague can easily revise or rebuild the artefact.

Table 1. The checklist applied to the exercise in Figures 1 to 4.

No.	Item	In the exercise
1	Level and aim in the prompt	Level (A1) and aim stated; no reason given for A1
2	How answers are judged	Stated for steps 1 to 4, not for step 5 (see Figure 4)
3	Trying it with answers that should fail	Partly done by the author: wrong and odd answers were tried (Figure 4); corrections were requested in the same conversation
4	A second look, where possible	Not done
5	Keeping the prompt	Prompt kept (Figure 1); other records not documented

Note. Entries record what the author did in writing the prompt and trying the exercise.

Table 1 applies the checklist to the exercise. The prompt settled the level, the aim, and how steps 1 to 4 judge an answer, but gave no reason for A1, and some faults began in the prompt itself: both the school item and the labels that an A1 learner may struggle with trace back to it. The prompt said nothing about how step 5 should judge a learner's sentence. Thus, the page applied a judgement the teacher had not specified, which became visible only when a sentence that should have failed was accepted (Figure 4). A trial reveals a wrong judgement only when the tester happens to enter the answer concerned, so an item that marks a defensible answer wrong, as in steps 2 and 3, can pass unnoticed in one person's trial, which is why item 4 suggests a second look. Ordinary trial use also cannot establish where learner data go or whether the page works for a learner using a screen reader.

The author checked the exercise by working through the page and did not read the code. Figures 2 to 4 show the version that the assistant first returned. The author then asked for corrections in the same conversation and used the revised exercise with learners.

The checklist rests on pedagogical judgement that teacher education should help teachers develop, and assessment of this work should give weight to the answers tried and the faults found, not to a page that runs.

Conclusion

Separating building from judging clarifies what AI-assisted coding changes for teachers. Building can be easier for some small artefacts when an assistant supplies code from a teacher's description. Judging whether the artefact is pedagogically and technically adequate still requires expertise and effort, and its standards do not depend on how the artefact was produced. Easier production does not justify lowering them, and the fact that an artefact works can conceal a failure to meet them.

Easier production can shift responsibility for materials design towards teachers. Decisions once settled by a publisher, a platform, or an instructional designer can now fall to the teacher who describes the artefact. They do not all arise at the same point: the order of the steps is fixed early, accessibility and data handling may be left to defaults unless someone raises them, and maintenance may remain open. The typology shows where these decisions sit for different kinds of artefact, and the account of costs shows what can follow when nobody takes those decisions.

Teaching future teachers to write code so that they can produce artefacts would prepare them mainly for the part of the work that has become easier. They need practice in working through the checklist proposed here: writing the prompt, trying the artefact before learners use it, and noting afterwards what learners did with it. Questions that cannot be settled by trying the page may require technical review, and artefacts kept in use over time also need someone responsible for their maintenance.

Most of the evidence drawn on here comes from English language teaching and from general software practice. The case for teacher-built materials is particularly strong in languages whose materials market is thin, and the typology needs testing against what teachers of those languages actually build. The checklist still needs to be evaluated in teacher education, both for its feasibility and for whether it helps teachers identify consequential design problems. Other questions concern how long teacher-built artefacts remain in use, what they contribute to learning, and, where an

artefact calls on a model each time it is used, how reliably the model handles the target language, particularly where evidence about its performance in that language is limited. An assistant could be designed to ask its user for the level, the objective, and the expected learner product, but what teachers would need to agree on before it could check those answers remains open. Whether a particular teacher should build a tool remains a professional judgement. Ease of production does not settle that question. Teachers must consider whether they or their institutions can adequately judge what they produce.

Declaration on the Use of Artificial Intelligence

Claude Opus 5 (Anthropic) produced the exercise shown in Figures 2 to 4 from the prompt reproduced in Figure 1. Claude Opus 5 was also used to assist with language revision and reference checks. The author reviewed all content and takes full responsibility for it.

References

Avelino, G., Passos, L., Hora, A., & Valente, M. T. (2016). A novel approach for estimating truck factors. In *2016 IEEE 24th International Conference on Program Comprehension (ICPC)* (pp. 1–10). IEEE. https://doi.org/10.1109/ICPC.2016.7503718

Barcomb, M., & Cardoso, W. (2025). Teachers as CALL customizers: Exploring teachers' perceptions of conceptualizing and customizing CALL materials. *Computer Assisted Language Learning, 38*(8), 2112–2138. https://doi.org/10.1080/09588221.2024.2414776

Becker, J., Rush, N., Barnes, E., & Rein, D. (2025). *Measuring the impact of early-2025 AI on experienced open-source developer productivity* (arXiv:2507.09089). arXiv. https://doi.org/10.48550/arXiv.2507.09089

Bouckaert, M. (2019). Current perspectives on teachers as materials developers: Why, what, and how? *RELC Journal, 50*(3), 439–456. https://doi.org/10.1177/0033688218810549

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934–957. https://doi.org/10.1080/09588221.2019.1648298

Deng, J., Fan, Z., & Meng, R. (2026). *Understanding the (in)security of vibe-coded applications* (arXiv:2606.23130). arXiv. https://doi.org/10.48550/arXiv.2606.23130

Dong, Y., Wang, S., & Li, Y. (2026). Enhancing human–AI communication in language education: Developing customized chatbots for teaching genre analysis and annotation. *Linguistics and Education, 93*, Article 101542. https://doi.org/10.1016/j.linged.2026.101542

Fu, Y., Liang, P., Tahir, A., Li, Z., Shahin, M., Yu, J., & Chen, J. (2025). Security weaknesses of Copilot-generated code in GitHub projects: An empirical study. *ACM Transactions on Software Engineering and Methodology, 34*(8), 1–34. https://doi.org/10.1145/3716848

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly, 15*(3), 219–236. https://doi.org/10.1080/15434303.2018.1453816

Guriță, A.-E., & Vatavu, R.-D. (2025). When LLM-generated code perpetuates user interface accessibility barriers, how can we break the cycle? In *Proceedings of the 22nd International Web for All Conference* (pp. 124–134). Association for Computing Machinery. https://doi.org/10.1145/3744257.3744266

Harsch, C., & Hartig, J. (2015). What are we aligning tests to when we report test alignment to the CEFR? *Language Assessment Quarterly, 12*(4), 333–362. https://doi.org/10.1080/15434303.2015.1092545

Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. *Instructional Science, 43*(2), 173–179. https://doi.org/10.1007/s11251-014-9343-4

Kennedy, C., & Levy, M. (2009). Sustainability and computer-assisted language learning: Factors for success in a context of change. *Computer Assisted Language Learning, 22*(5), 445–463. https://doi.org/10.1080/09588220903345218

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing, 16*(2), 81–96. https://doi.org/10.1016/j.asw.2011.02.003

Ko, A. J., Abraham, R., Beckwith, L., Blackwell, A., Burnett, M., Erwig, M., Scaffidi, C., Lawrance, J., Lieberman, H., Myers, B., Rosson, M. B., Rothermel, G., Shaw, M., & Wiedenbeck, S. (2011). The state of the art in end-user software engineering. *ACM Computing Surveys, 43*(3), Article 21. https://doi.org/10.1145/1922649.1922658

Kranch, D. A. (2008). Who owns online course intellectual property? *Quarterly Review of Distance Education, 9*(4), 349–356. https://doi.org/10.1108/qrde-12-2008-0002

Lieberman, H., Paternò, F., Klann, M., & Wulf, V. (2006). End-user development: An emerging paradigm. In H. Lieberman, F. Paternò, & V. Wulf (Eds.), *End user development* (pp. 1–8). Springer. https://doi.org/10.1007/1-4020-5386-X_1

Lo, A. W. T. (2025). The educational affordances and challenges of generative AI in Global Englishes-oriented materials development and implementation: A critical ecological perspective. *System, 130*, Article 103610. https://doi.org/10.1016/j.system.2025.103610

McKenney, S., Kali, Y., Markauskaite, L., & Voogt, J. (2015). Teacher design knowledge for technology enhanced learning: An ecological framework for investigating assets and needs. *Instructional Science, 43*(2), 181–202. https://doi.org/10.1007/s11251-014-9337-2

Mishan, F. (2022). The global ELT coursebook: A case of Cinderella's slipper? *Language Teaching, 55*(4), 490–505. https://doi.org/10.1017/S0261444820000646

Mørch, A. I., Ludvigsen, S., & Gilje, Ø. (2025). Teachers as end-user developers: Two case studies of adapting language models for education. In V. Paneva, D. Tetteroo, V. Frau, S. Feger, D. Spano, F. Paternò, S. Sauer, & M. Manca (Eds.), *Joint proceedings of the workshops, work in progress demos and doctoral consortium at the IS-EUD 2025* (CEUR Workshop Proceedings, Vol. 3978). https://ceur-ws.org/Vol-3978/short-s2-07.pdf

Otto, S. E. K., & Pusack, J. P. (2009). Computer-assisted language learning authoring issues. *The Modern Language Journal, 93*(s1), 784–801. https://doi.org/10.1111/j.1540-4781.2009.00973.x

Özçelik, Ö., & Kennedy Kent, A. (2023). *Designing effective language learning materials for less commonly taught languages: A research-based guide.* Georgetown University Press.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The impact of AI on developer productivity: Evidence from GitHub Copilot* (arXiv:2302.06590). arXiv. https://doi.org/10.48550/arXiv.2302.06590

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sarı, Y. E. (2025). Üretken yapay zekâ ile beceri temelli Almanca eğitimi [Skill-based German language instruction with generative artificial intelligence]. In Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (pp. 230–251). Günce Yayınları.

Sarkar, A., & Drosos, I. (2025). Vibe coding: Programming through conversation with artificial intelligence. In *Proceedings of the 36th Annual Workshop of the Psychology of Programming Interest Group (PPIG 2025)* (pp. 87–118). Psychology of Programming Interest Group. https://www.ppig.org/files/2025-PPIG-36th-sarkar.pdf

Shin, D., Lee, J. H., Lee, E., & Lee, Y. (2026). No-code approaches to developing ELT resources with generative artificial intelligence. *ELT Journal, 80*(1), 164–173. https://doi.org/10.1093/elt/ccaf050

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching, 45*(2), 143–179. https://doi.org/10.1017/S0261444811000528

Tomlinson, B. (2023). Principled blended learning materials for language acquisition. *Journal of English Language Teaching Innovations and Materials (Jeltim), 5*(1), 1–16. https://doi.org/10.26418/jeltim.v5i1.60159

U.S. Copyright Office. (2025). *Copyright and artificial intelligence, Part 2: Copyrightability.* Library of Congress. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf

Wiley, D., & Hilton, J. L., III. (2018). Defining OER-enabled pedagogy. *International Review of Research in Open and Distributed Learning, 19*(4), 133–147. https://doi.org/10.19173/irrodl.v19i4.3601

Wong, M. H. I., Zhao, H., & MacWhinney, B. (2018). A cognitive linguistics application for second language pedagogy: The English preposition tutor. *Language Learning, 68*(2), 438–468. https://doi.org/10.1111/lang.12278

Yu, Z., Wu, Y., Zhang, N., Wang, C., Vorobeychik, Y., & Xiao, C. (2023). CodeIP-Prompt: Intellectual property infringement assessment of code language models. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 202, pp. 40373–40389). PMLR. https://proceedings.mlr.press/v202/yu23g.html

Zhao, H., Huang, S., Zhou, Y., & Wang, R. (2020). Schematic diagrams in second language learning of English prepositions: A behavioral and event-related potential study. *Studies in Second Language Acquisition, 42*(4), 721–748. https://doi.org/10.1017/S027226311900069X

Reconceptualizing EFL Pedagogy in the Age of Generative AI and Digitalization

Aysel Deregözü [1] *& İlayda Erdem* [2]

Abstract
The shift to post-method pedagogy liberated language teaching from rigid, one-size-fits-all methodologies, putting the emphasis instead on context-sensitive, autonomous, and reflective pedagogical practices. Simultaneously, the rapid rise of digital technologies, particularly generative Artificial Intelligence (GenAI), has reshaped the modern foreign language learning landscape. This chapter examines the conceptual relationship between post-method principles and AI-mediated learning environments. Specifically, it explores how generative AI tools operationalize key post-methodological macro-level strategies, such as maximizing learning opportunities, supporting learner autonomy, and facilitating context-based language exposure. Tracing the shift from Computer-Assisted Language Learning (CALL) applications to dynamic AI environments, the discussion highlights how adaptive systems allow for individualized learning while bringing up serious questions about data bias, digital equity, and pedagogical integrity. Rather than defining GenAI as a replacement for previous methodological tenets, this chapter proposes AI as an adaptive scaffolding mechanism that allows educators and learners to co-create responsive, contextually grounded educational settings. The chapter concludes with actionable insights for teacher education, ethical engagement, and future empirical inquiry.
Key Words: Post-Method Pedagogy, Generative AI, Digitalization in ELT, AI-Assisted Language Learning (AIALL), Adaptive Learning Environments

Introduction

Over time, several approaches have been tried and major paradigm shifts have taken place as an offshoot of changing credences regarding effective language teaching (Bacus, 2021). During the last ten decades, the search for a more effective way of teaching a second or foreign language went through numerous innovations (Bhandari & Bhandari, 2025). Initially dominated by teacher-centered paradigms like the Grammar-Translation Method (GTM), which arose to focus on classical Latin and Greek texts, linguistic forms, and literary translation (Hilgendorf, 2020), foreign

[1] Assoc. Prof. Dr., İnönü University, Faculty of Education, Department of Foreign Language Education, English Language Teaching Program, ORCID: 0000-0002-8371-8047, *aysel.deregozu@inonu.edu.tr*, Türkiye

[2] Res. Asst., Fırat University, Faculty of Education, Department of Foreign Language Education, English Language Teaching Program, ORCID: 0009-0005-0721-2393, *ierdem@firat.edu.tr*, Türkiye

language instruction was heavily criticized for not emphasizing real-life communicative contexts (Aslan, 2025). Subsequently, practitioners developed a range of methods including Audiolingualism, the Direct Method, the Natural Approach, the Silent Way, and Total Physical Response. The introduction of Communicative Language Teaching (CLT) in the 1970s, along with its descendants such as Content-Based Instruction (CBI) and Task-Based Language Teaching (TBLT), triggered a period of rapid methodological expansion (Karimvand et al., 2014).

However, by the late 1980s and 1990s, English language teaching (ELT) practitioners, instructors, and theorists started to express unhappiness as the core assumptions of method-based instruction were questioned (Senen, 2024). Conventional methods were found to be highly hypothetical, built on crumbled theoretical ground, and untouched by practicality. Prescribed methods were too restricted to explain the complexity of teaching and largely ignored the institutional, sociopolitical, and contextual conditions within which teachers work (Kumaravadivelu, 2006).

This dissatisfaction led directly to the emergence of the post-method pedagogy, introduced by Kumaravadivelu in 1994 as a turning point to overcome the limitations of rigid, one-size-fits-all methodologies (Bhandari & Bhandari, 2025). Post-method pedagogy represents a departure from the century-old concept of method, accepting multiple truths and options in English language teaching (Bacus, 2021). It posits that all methods are context-specific and encourages teachers to adjust, adapt, and alter practices based on classroom realities. Under this paradigm, teachers transition from method-followers into decisive elements who construct and deconstruct classroom systems based on their own knowledge, logic, and experiences (Bacus, 2021). The post-method framework addresses the visualization of a three-dimensional structure that encompasses the parameters of *particularity, practicality*, and *possibility* (Kumaravadivelu, 2001). A post-method approach to pedagogy must accomplish three things: first, it should support the development of language teaching that is genuinely responsive to the specific linguistic, cultural, and political conditions of a given context, rather than applying one-size-fits-all methods. Second, it should break down the traditional hierarchy separating theorists from practitioners, empowering teachers to develop their own personal theories grounded in their classroom experience. Third, it should recognize and draw on the sociopolitical awareness that learners and teachers bring with them, using it to support their ongoing search for identity and to contribute to broader social change (Kumaravadivelu, 2001).

To operationalize these parameters in classroom practice, Kumaravadivelu (2003a) designed a strategic framework consisting of ten macro-

strategies such as maximizing learning opportunities, fostering learner autonomy, facilitating negotiated interaction, contextualizing linguistic input, and fostering language awareness, designed to sharpen teachers' sensitivity to what happens in the classroom. This heightened awareness, in turn, allows teachers to devise diverse, context-appropriate instructional approaches tailored to their specific teaching situations (Karimvand et al., 2014).

Unlike traditional micro-strategies or fixed techniques, these macro-strategies function as broad instructional guidelines within which practitioners can generate location-specific classroom activities. Kumaravadivelu (2003b) regards these parameters "as the axle that connects and holds the center of the pedagogic wheel" (p. 41) and the macrostrategies "as spokes that join the pedagogic wheel to its center thereby giving the wheel its stability and strength" (p. 41).

In parallel with these theoretical shifts, dramatic changes have taken place due to the rapid integration of digital technologies in language education. The field has evolved from early audio-visual tools to advanced digital applications, online platforms, and mobile devices (Banafi, 2025; Martinaj, 2025; Moradi, 2025; YarAhmadi & Kargar Behbahani, 2025). The increasing integration of Artificial Intelligence (AI), particularly generative AI, has garnered significant academic interest. By generating text, speech, and multimodal media from data, Generative Artificial Intelligence (GenAI) has the potential to fundamentally transform pedagogical approaches in English as a foreign language (EFL) education, offering adaptive, context-responsive outputs that enhance tools for both instructors and learners (Lim et al., 2023).

Despite these promising advancements, literature reveals a tension between the pedagogical potential of GenAI and the challenges of its implementation. A significant research gap remains regarding the practical adoption of GenAI by lecturers for EFL preparation. Concerns related to reliability, ethical awareness, overreliance, academic integrity, and algorithmic bias underscore the need for structured integration (Peng, 2026). Understanding the opportunities and challenges of AI in this context is essential to enhancing instructional efficacy.

The purpose of this chapter is to examine the conceptual relationship between post-method principles and AI-mediated learning environments. Rather than defining GenAI as a new prescriptive method to replace previous tenets, this chapter proposes AI as an adaptive scaffolding mechanism that empowers teachers and learners to co-create flexible, contextually grounded educational settings.

Theoretical Background

The Development of Post-Method Pedagogy

In the early and mid-20th century, ELT researchers proposed various methods as effective frameworks for language teaching. The Audio-Lingual Method (ALM), which is based on the behaviorist school of thought, has prioritized accuracy and rote learning through repetition, while placing less focus on the meaningful usage of language in different contexts (Nepal, 2023). Similarly, originating from the study of ancient Greek and Latin literature, the Grammar-Translation Method and similar classical approaches concentrated on isolated grammar rules and their accurate instruction (Aydın, 2025), primarily aiming to translate written literary texts (Hilgendorf, 2020).While the 1970s brought methods focused on oral proficiency and real-life communication—such as CLT, the Natural Approach, Total Physical Response, and the Silent Way—these frameworks continued to prescribe teaching procedures globally without considering local, practical, and political factors (Senen, 2024).

By the 1990s, theorists and practitioners voiced displeasure with these restrictions, giving rise to the post-method era. Scholars asserted that prescribed methods with advanced principles may not be suitable or effective for all teaching and learning contexts (Bhandari & Bhandari, 2025). Method-based models overlooked the necessity for language instruction to be tailored to the individual cohort of educators teaching a distinct group of learners with particular objectives within a defined institutional and sociocultural context (Senen, 2024).

Kumaravadivelu (1994, 2001, 2003a) developed post-method pedagogy as an alternative to methods rather than another method. Post-method pedagogy looks at instructional resources for real-world communication in second language (L2) classrooms, encouraging students to become fluent and accurate language users as participants in a cooperative endeavor.

The post-method framework is anchored by three foundational parameters (Kumaravadivelu (2001; 2003a):

- **Particularity:** Emphasizes local context and "situational understanding". It is a bottom-up approach that respects local linguistic, socio-cultural, and political features. Policymakers and L2 managers must recognize local circumstances to support efficient instruction.
- **Practicality:** Focuses on empowering teachers to formulate personal theories based on classroom experiences, bridging the gap between theory and practice.

- **Possibility:** Fosters critical consciousness, acknowledging that language teaching involves socioeconomic status, political realities, and struggle between old and new identities.

Using a wheel metaphor, Kumaravadivelu (2003b) positioned the three parameters at the central axle, connected and strengthened by the 10 macro-strategies that serve as supporting spokes.

Digital environments exhibit a natural alignment with post-method principles. Post-method language education no longer revolves around finding a single "good method," but focuses on developing class assignments and activities that match second language acquisition research and suit the specific characteristics of the class. Digital platforms provide customized strategies and flexibility, allowing teachers and students to draw on their own experiences, values, and cultures (Bhandari & Bhandari, 2025). By moving away from standardized textbook constraints, digital technology provides the infrastructure needed to realize context-sensitive, autonomous L2 learning.

Digital environments allow learners to explore personalized learning paths, select materials aligned with their personal interests, and engage in autonomous practice outside formal classroom hours (Brugliera, 2024). Furthermore, digital tools allow teachers to curate, adapt, and co-create instructional materials tailored to their specific classroom demographics (Simos & Katsiampoura, 2025), operationalizing the principle of *particularity*. Rather than forcing students into a uniform learning sequence, digital platforms provide the flexibility required to accommodate individual learning rates (Bahari et al., 2025). Thus, the shift toward post-method pedagogy pairs naturally with digital technologies for implementation—a relationship that has grown even stronger with the emergence of generative AI.

The Age of Digitalization in English Language Education

The trajectory of technology integration in ELT over the past six decades reflects broader developments in computer science, educational technology, and learning theory. This evolution can be understood as a three-stage historical progression: Computer-Assisted Language Learning (CALL), Mobile-Assisted Language Learning (MALL), and AI-Assisted Language Learning (AIALL).

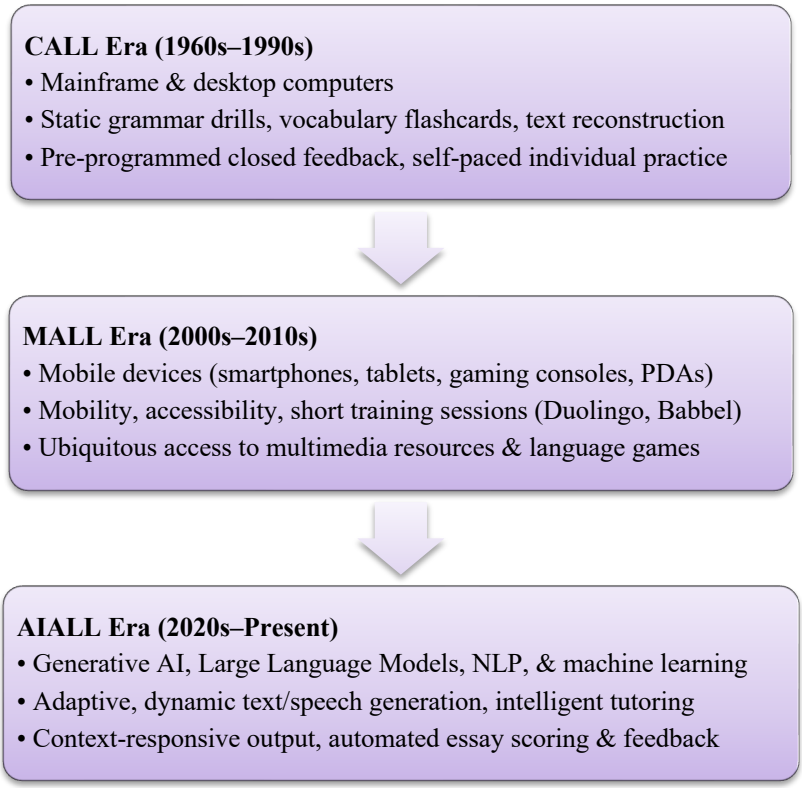


Fig. 1. Evolution of Digital Language Pedagogy: From Traditional CALL to AIALL

The earliest iteration, Computer-Assisted Language Learning (CALL), emerged in the 1960s and evolved through structural, communicative, and integrative phases (Aslan, 2025; Warschauer & Healey, 1998). Technological limitations initially forced CALL exercises to mirror Grammar-Translation and Audiolingual frameworks, concentrating heavily on decontextualized grammar practice, flashcards, and text translation (Babayiğit & Çoşkun, 2023). By the late 1990s, integrative CALL utilized multimedia technology and the early Internet to connect language learning with authentic communicative contexts, providing exposure to authentic input, diverse accents, and multimedia resources (Nazeer et al., 2023). Studies confirmed that CALL promoted learner autonomy by enabling self-paced, independent learning outside the classroom (; Jeon et al., 2023; Liu & Gablasova, 2023; Xiao et al., 2023). However, despite these advances, early CALL applications remained constrained by desktop

hardware, static content architecture, and rule-based software that provided pre-programmed, closed-ended feedback.

Mobile-Assisted Language Learning (MALL) emerged with the widespread adoption of mobile devices such as smartphones, tablets, PDAs, MP3 players, and electronic dictionaries (Stockwell, 2022). MALL liberated language learning from physical computer labs, enabling ubiquitous, contextualized, and informal learning experiences. Applications such as Duolingo, Babbel, and HelloTalk transformed language study by providing convenient, short daily training sessions accessible via mobile devices. MALL enabled practice of vocabulary, grammar, and pronunciation in short available periods, fostering engagement through gamified tasks (Aslan, 2025). However, despite their portability, many MALL applications still relied on pre-scripted content sequences and automated multiple-choice scoring systems, limiting the depth of interactive language negotiation.

Artificial Intelligence-Assisted Language Learning (AIALL) represents a qualitative shift from earlier CALL and MALL technologies. While earlier CALL tools delivered pre-programmed content, AIALL leverages Natural Language Processing (NLP), machine learning, and Large Language Models (LLMs) to create adaptive learning environments and intelligent tutoring systems (Alqahtani et al., 2023; Kamalov et al., 2023).

GenAI tools (such as ChatGPT, Claude or Gemini) create tailored, multimodal responses, including written content, visual assets, audio, and dynamic simulations, in response to natural language prompts (Lim et al., 2023). Unlike earlier rule-based tools, generative AI creates original language in context, challenging conventional boundaries between learner, teacher, and tool (Kasneci et al., 2023; Kozlova, 2026).

GenAI applications have been shown to support language teachers across a range of instructional functions. In terms of instructional preparation and material design, GenAI assists teachers with lesson planning, material development, and assessment preparation, while also enabling innovative practices such as game-based activities and data-driven learning approaches (Kartal, 2024; Kusuma et al., 2024; Moorhouse et al., 2024; Raimkulova et al., 2025). Beyond preparation, GenAI has also proven valuable in the drafting, revision, and feedback process, where its use has been linked to gains in improvements in textual structure, grammatical precision, and vocabulary growth (Athanassopoulos et al., 2023; Guo, 2025); notably, contemporary Natural Language Processing (NLP) models can now evaluate coherence, vocabulary, and grammatical accuracy at levels that approach human inter-rater agreement (Hoffmann, 2026). In addition to supporting writing development, conversational chatbots and

simulations, such as ELSA, Google Assistant, replicate natural dialogue, offering learners opportunities for interactive speaking practice alongside real-time feedback on both spoken and written output (Yang et al., 2022). Finally, GenAI's capacity to deliver immediate, adaptive feedback has been associated with important affective benefits, including reduced language anxiety relative to traditional classroom settings, as well as greater learner enjoyment and confidence (Hoang, 2026; Liu et al., 2024).

These developments illustrate a contrast between personalized learning pathways and standardized method constraints. Traditional methods enforced a single coursebook, one instructor, and rigid sequences (Aslan, 2025). Every student in a given classroom was expected to move through identical textbook units, practice identical dialogic scripts, and complete identical assessment tasks. This structural rigidity assumes a homogenous learning pace and uniform learner background.

In contrast, AIALL dynamically adapts content difficulty, schedules multimodal vocabulary presentation (combining visual, auditory, and contextual modalities), and tracks longitudinal performance (Luo, 2026). By tailoring instruction to individual preferences and goals, GenAI shifts technology from a simple delivery mechanism to a multifunctional pedagogical resource (Peng, 2026).

This technological flexibility removes the structural requirement for a one-size-fits-all curriculum. Instead of fitting the student to a rigid instructional method, generative AI tools allow language learning to adapt to the learner's specific background, goals, pace, and cultural context. In this way, AIALL provides a technological infrastructure capable of realizing the post-method ideal of personalized, context-sensitive language education.

Integrating Artificial Intelligence with Post-Method Macro-Strategies

To evaluate the pedagogical value of Generative AI without falling into the trap of technological determinism, AI integration must be guided by a clear theoretical framework. Kumaravadivelu's (1994, 2001, 2003a) post-method macro-strategies offer an effective framework for this purpose. Rather than serving as another set of prescriptive rules, these macro-strategies function as dynamic tools that empower teachers to construct classroom practices based on their own professional autonomy, local conditions, and student needs (Kumaravadivelu, 2003a). Simultaneously, the emergence of generative artificial intelligence and Artificial Intelligence-

Assisted Language Learning (AIALL) has introduced flexible, multimodal, and context-responsive capabilities into EFL education (Yiling et al., 2025).

Unlike earlier Computer-Assisted Language Learning (CALL) tools that delivered pre-programmed content through static drills, modern GenAI tools generate original language, speech, images, and interactive simulations in real time (Kozlova, 2026). Far from functioning merely as automated substitutes for traditional instruction, AI-driven technologies can operationalize key post-method macro-strategies. By serving as a multifunctional pedagogical resource, GenAI offers dynamic scaffolding that allows educators and learners to realize post-method principles in practice.

Macro-Strategy 1: Maximizing Learning Opportunities

Kumaravadivelu's (2003a) macro-strategy of maximizing learning opportunities emphasizes that teachers must create and utilize learning opportunities both inside and outside the classroom, treating learners as active participants in a cooperative endeavor. In traditional EFL environments, however, learning opportunities are often constrained by fixed classroom schedules, single coursebooks, and limited individual interaction time with the instructor (Zhang et al., 2019). GenAI tools offer a viable solution to these structural limitations. Serving as scalable, low-anxiety extensions of the classroom environment (Peng, 2026), AI-driven language platforms transcend conventional spatial and temporal boundaries by providing round-the-clock linguistic assistance (Song et al., 2022). Learners no longer have to wait for classroom contact hours or teacher feedback; instead, they can access self-paced, individualized learning experiences tailored to their specific proficiency levels and preferences whenever needed (Samsudin & Allen, 2026).

Voice-enabled conversational AI platforms are designed to mimic natural dialogue, processing what learners say and producing responses that are both contextually appropriate and meaningful (Jeon, 2022; Lee & Jeon, 2022). Some more sophisticated AI systems built into language learning apps, such as ELSA or Praktika, go further, allowing students to engage in simulated conversations with a virtual partner (Aslan, 2025). Such interactions promote social engagement and collaborative learning, consistent with sociocultural theory's emphasis on the central role of social interaction in acquiring language (Tai & Chen, 2023). By rehearsing language skills in realistic scenarios, getting instant feedback, and taking part in genuine enhances, learners are able to strengthen their overall communicative ability (Bahari et al., 2025).

Furthermore, foreign language anxiety is a well-documented barrier, often compounding general social anxiety; students dread negative evaluation

during oral production or writing tasks (Horwitz et al., 1986). In this realm, GAI tools function as highly appealing, low-anxiety interactive partners, allowing students to practice target languages without fear of judgment. By allowing students to experiment with the structures of the target language, make mistakes, and ask for real-time explanations without fear of negative social judgment from their peers or instructors, these GAI tools can reduce anxiety about speaking the language and, as a result, boost self-esteem and autonomy (Zou & Wang, 2024).

Macro-Strategy 2: Fostering Learner Autonomy

Learner autonomy, described as the capacity to take charge of one's own learning journey, is central to post-method pedagogy (Holec, 1981; Little, 1991). In an AIALL environment, autonomy evolves from individual study worksheets to a dynamic, self-directed feedback loop (Han, 2025).

As learners interact with AI in a more self-reliant way, the concept of autonomy itself starts to shift. Students gain more say over the timing and manner of their learning, able to draw on resources without needing to wait for a teacher's guidance. Empirical studies support this shift, demonstrating that AI use actively promotes independent learning and personal discovery (Salah et al., 2024; Yalçınalp et al., 2024). A prime example of this dynamic is prompt engineering, through which learners take on an active role in managing their own learning process, submitting a draft essay to an LLM for targeted feedback rather than waiting for the teachers' periodic assessment (Kundu & Bej, 2025). This mode of engagement shifts the learner from a passive receiver of corrections to an active agent involved in metacognitive reflection. Learners can independently adjust the difficulty of study materials, seek targeted grammar clarifications, design personalized vocabulary lists, and self-assess their progress, cultivating skills essential for lifelong language learning (Wei, 2023).

Learner autonomy is central to post-method pedagogy, which seeks to empower students to take control of their learning process, make independent decisions, and develop self-directed learning strategies (Kumaravadivelu, 2001) AI tools enhance student autonomy through two primary frameworks (Owoseni et al. 2024):

- **AI-Enabled Learning:** Supports self-paced, independent instruction where learners regulate their own study schedules, set personal goals, and complete tasks with minimal teacher mediation.
- **AI-Supported Learning:** Provides adaptive scaffolding, automated diagnostics, and immediate feedback that guide students during self-directed study.

Generative AI further expands learner agency by offering choice-driven feedback rather than rigid enforcement. For example, platforms like Grammarly offer suggestions that users can evaluate, accept, or reject, allowing them to retain control over their final writing choices (Yang, 2026). Conversational agents like ChatGPT act as guides during brainstorming and drafting, scaffolding student ideas without prescribing fixed outcomes (Deep et al., 2025; Xiao & Zhi, 2023). By serving as continuous 24/7 learning assistants, AI tools shift learners from passive receivers of instruction to active decision-makers who direct their own revision, vocabulary exploration, and communicative practice (Chen & Lin 2023; Kundu & Bej 2025).

Macro-Strategy 3: Facilitating Negotiated Interaction

Facilitating negotiated interaction requires environments where learners actively participate in meaningful, context-responsive communication, testing hypotheses and negotiating meaning rather than performing mechanical repetition (Kumaravadivelu, 2003a). Traditional methods like the Audio-Lingual Method relied on static pattern drills, providing little space for genuine, negotiated communication (Nepal, 2023).

Advanced AIALL applications, including conversational chatbots, virtual language partners, and intelligent tutoring systems, facilitate negotiated interaction by simulating real-world dialogue and communicative tasks (Lim et al., 2023) Systems such as ELSA, ChatGPT, or Google Assistant allow learners to engage in responsive, interactive scenarios such as ordering food in a restaurant, negotiating a business transaction, or engaging in open-ended conversation (Aslan, 2025).

Crucially, AI-mediated interactive environments create a safe, non-judgmental learning space that significantly reduces language-speaking anxiety and the fear of negative social evaluation (Liu et al., 2024). By removing affective barriers, AI tools encourage learners to take risks, test novel expressions, build oral fluency, and sustain communicative interaction over extended periods, thereby promoting deeper language processing (Bhar, 2026).

Macro-Strategy 4: Contextualizing Linguistic Input

Post-method pedagogy highlights the parameter of particularity, asserting that effective language learning must be firmly rooted in local contexts, situations, and the practical surroundings of the learners (Kumaravadivelu, 2001). Traditional textbooks frequently present target language forms in decontextualized, highly artificial structures that fail to connect with students' immediate environments (Zhang et al., 2019).

AI technologies provide novel mechanisms for connecting target language input directly to authentic, real-world surroundings through dynamic contextualization (Hwang et al., 2023). Integrating AI with recognition-to-text technologies, such as Speech-to-Text Recognition (STR) and Image-to-Text Recognition (ITR), enables learners to capture elements from their immediate environment (e.g., taking photos of objects or dictating local events) (Shadiev et al., 2018; Shadiev et al., 2020). Natural Language Processing (NLP) algorithms can then instantly generate contextually relevant target vocabulary, sample sentences, and contextual writing prompts based on those local inputs (Du et al., 2021; Hwang et al., 2023; Rees, 2022).

Furthermore, large language models adjust linguistic input to match SLA criteria for optimal input, ensuring high frequency, salience, comprehensibility, and interactional relevance (Zou et al., 2023). By tailoring complexity and thematic context to the specific situation of a given group of learners, AI allows language instruction to remain context-sensitive and locally relevant (Aslan, 2025).

Macro-Strategy 5: Fostering Language Awareness

Fostering language awareness involves encouraging learners to attend to the formal, functional, and structural properties of the target language during meaningful use (Kumaravadivelu, 2003a). Rather than teaching grammar rules through isolated drills, post-method practice encourages explicit reflection on language forms as they arise in context (Bhandari & Bhandari, 2025).

Contemporary AI tools leverage advanced NLP architectures to provide immediate, individualized feedback on learner language production, addressing grammatical accuracy, lexical diversity, structural coherence, and pronunciation (Hoffmann, 2026). Speech-to-Text Recognition tools transcribe spoken input (Shadiev et al., 2018), allowing learners to visually analyze their own oral production, pinpoint phonological errors, and perform self-correction (Dai &Wu, 2021).

Rather than offering simple error correction, AI-driven writing tools can evaluate text organization, suggest sentence reorganizations (Alm, 2024; Sajja et al., 2026), propose diverse vocabulary options based on the original writing (Wang, 2020), and prompt learners to reflect on their linguistic choices (Yang, 2026). This automated feedback stimulates metacognitive reflection, helping learners compare their output with AI-generated suggestions, analyze communication strategies, and refine their overall language awareness (Creely, 2024).

However, because AI feedback systems may struggle with subtle pragmatic intent, cultural nuance, and deep meaning-making, automated tools should be understood as complementary mechanisms (Bhar, 2026). They work best when combined with human-mediated evaluation and reflective teacher guidance to ensure comprehensive language development (Torres & Kahveci, 2025).

Challenges and Ethical Considerations

While the integration of GenAI into post-method EFL instruction offers flexible, context-sensitive learning environments, it simultaneously introduces substantial pedagogical, technical, and sociopolitical challenges (Creely, 2024). Gaining insight into both the difficulties and opportunities that come with using AI in this context could meaningfully improve teaching effectiveness and student success (Chiu, 2024; Jiang, 2022).

The Digital Divide and Unequal Access

The application of post-method principles, specifically the parameter of *particularity*, demands high sensitivity to local socioeconomic, cultural, and political realities (Kumaravadivelu, 2003a). However, the deployment of educational AI tools often reinforces systemic inequities through multi-tiered digital divides and embedded biases (Chinta et al., 2024).

A pronounced gap persists between well-resourced institutions and socioeconomically disadvantaged or developing regions (Conceição & Van Der Stappen, 2025; Druga et al., 2022). Restricted affordability (Kim, 2024; Song et al., 2022), inadequate technological infrastructure, and limited internet connectivity severely constrain equal access to advanced AI platforms, directly undermining educational equity (Vesna et al., 2025).

Beyond access alone, differential quality of AI output produces uneven educational outcomes. Students in under-resourced communities are more likely to encounter lower-quality tools or less supportive feedback systems, compounding existing historical disadvantages (Bulathwela et al., 2024). As a result, the indiscriminate implementation of educational AI widens an existing developmental discrepancy, further marginalizing underprivileged students in the global ELT landscape.

Algorithmic Bias and Cultural Hegemony

Most dominant GenAI models are trained on datasets predominantly derived from Western, native-English-speaking populations (Ferrara, 2023). Training AI on these extensive databases carries the distinct hazard of yielding linguistic outputs that neglect the intrinsic pluralism, local

variants, and dialectal nuances of human language (Creely, 2024). As a result, these systems frequently lack recognition of indigenous languages, local cultural contexts, and non-native English varieties (Mac Fadden et al., 2024). This linguistic and cultural bias reduces the pedagogical relevance of AI systems for diverse learner populations, effectively reproducing exclusion through technology design rather than mere physical access—a manifestation conceptualized as the "third-level digital divide" (Matjie et al., 2026). When implemented without critical oversight, these algorithmic systems misinterpret diverse student work, generate culturally inappropriate feedback, enforce a homogenized, sanitized version of the target language, and systematically marginalize underrepresented learner populations (Creely, 2024; Matjie et al., 2026).

Beyond the immediate issues of representation, the proliferation of biased GenAI outputs creates a subtle yet pervasive form of cultural hegemony within ELT. Traditional language resources have long been criticized for adopting a "one-size-fits-all" model, but the adaptive and conversational nature of GenAI creates an illusion of personalized neutrality while secretly reinforcing dominant Western socio-cultural norms (Creely, 2024). By presenting standardized native-speaker norms as the universal benchmark for accuracy and pragmatic appropriateness, AI-driven feedback risks alienating foreign language EFL learners from their own sociocultural identities. This dynamic stands in direct tension with the core tenets of *Post-Method Pedagogy* (Kumaravadivelu, 2006). According to the parameters of *particularity*, *practicality*, and *possibility*, effective language education must be context-sensitive, embedded in local sociopolitical realities, and empowering of learners' transformative potential (Bhandari & Bhandari, 2025; Kumaravadivelu, 2006). Algorithmic bias directly undermines the pedagogy of particularity and possibility by imposing top-down, Western-centric epistemologies under the guise of automated objectivity (Mac Fadden et al., 2024).

Over-Reliance and the Risk of Cognitive De-skilling

While AI tools support learner autonomy through personalized feedback, excessive dependency on automated assistance can inadvertently undermine cognitive awareness (Farhat, 2026) and deep learning (Haque & Li 2024).

When learners rely habitually on tools like ChatGPT to construct essays, generate ideas, or fix syntactic errors without reflection, they bypass essential cognitive struggle. Over time, this offloading weakens critical thinking (Haque & Li 2024), reduces attention spans (Fu & Hanaki 2026), and impairs unaided problem-solving capabilities (Noorbehbahani & Oyibo, 2026).

The continuous availability of low-anxiety, non-judgmental AI chat partners can lead learners to prefer virtual interactions over human communication (Epley et al. 2007; Wan & Chen 2021). This dynamic risks diminishing interpersonal competence, reduced empathy, and heightened anxiety whenever automated support is not available (Bonezzi et al. 2022; Marriott & Pitardi 2024). Ultimately, this double-edged reliance may undermine social confidence and accelerate the gradual de-skilling of core linguistic and cognitive competencies that are crucial for authentic, real-world communication.

Redefining the Practitioner's Identity

The integration of AI disrupts traditional pedagogical hierarchies, sparking tension surrounding teacher identity and classroom authority (Zaman et al., 2024). Post-method pedagogy highlights teachers as autonomous decision-makers who construct classroom practices based on their professional experience and local needs (Kumaravadivelu, 2001; 2003a). As AI systems take over material creation, immediate error detection, and dialogic coaching, teachers must transition from authoritative knowledge providers to co-navigators and facilitators of learning (Gentile et al., 2023).

Automated real-time feedback can create anxieties regarding professional obsolescence or marginalization (Namaziandost & Rezai, 2024; Han et al., 2025). However, empirical evidence shows that student motivation, emotional engagement, and meaningful language progress depend heavily on teacher presence, empathy, humor, and dialogic feedback (Derakhshan, 2025; Mahapatra, 2024). The educational worth of artificial intelligence is significantly influenced by the manner of its implementation, and the efficacy of AI-enhanced learning relies on its incorporation into human-centric teaching models (Holmes et al., 2019). AI cannot replicate the emotional, relational, and cultural guidance provided by human educators (Almashour et al., 2025).

Generative AI is a powerful pedagogical resource, but it is not a self-sustaining educational solution (Adewale et al., 2026). Maintaining the integrity of post-method language education requires balancing technological capabilities with critical human judgment (Peng, 2026), ethical oversight (Celik, 2023; Celik et al., 2026), and a steadfast commitment to context-sensitive instruction (Kumaravadivelu, 2003a; 2003b).

Conclusion and Future Directions

The integration of GenAI into English as a foreign language instruction represents a profound paradigm shift, redefining the boundaries between

language learners, educators, and digital tools (Banafi, 2025; Martinaj, 2025; Moradi, 2025; YarAhmadi & Kargar Behbahani, 2025). Rather than acting simply as fixed software or static information stores, GenAI tools function as flexible, context-aware, and adaptive resources for teaching (Sardi et al., 2025), blurring the traditional lines that separate learner, teacher, and tool (Kasneci et al., 2023). By facilitating real-time feedback (Hwang et al., 2023; Kundu & Bej, 2025; Yang, 2026; Zou et al., 2023), authentic dialogic simulations (Aslan, 2025; Yiling et al., 2025), personalized writing support (Hwang et al., 2023; Woo et al., 2024), and speech-to-text input recognition (Shadiev et al., 2018; Shadiev et al., 2020), AI-mediated instruction significantly enhances student autonomy, language proficiency, and interactional confidence (Ahmad & Khasawneh, 2023; Ding & Yusof, 2025; Han, 2025).

Crucially, this technological evolution aligns directly with the core tenets of post-method pedagogy. As language education transitions away from rigid, "one-size-fits-all" traditional methods toward context-sensitive instructional frameworks (Bhandari & Bhandari, 2025), GenAI offers practical avenues to operationalize the parameters of *particularity*, *practicality*, and *possibility.* Within this framework, particularity is realized as GenAI customizes learning materials (Amin, 2023) to address local, to level-appropriate linguistic, and cultural classroom realities (Bhar, 2026). Practicality manifests through empowering teachers to generate context-responsive personal theories (Lim et al., 2023) and dynamic classroom tasks rather than relying strictly on static coursebooks (Aslan, 2025). Furthermore, possibility is fostered by enhancing learner agency (Yang, 2026), critical reflection and self-directed exploration beyond prescriptive method boundaries (Kumaravadivelu, 2003b)

However, the pedagogical potential of GenAI is counterbalanced by significant challenges and systemic limitations (Kannan & Munday, 2018). Issues surrounding algorithmic and cultural bias, the multi-tiered digital divide, content reliability and hallucinations, cognitive overreliance, academic integrity, and shifts in teacher identity underscore that technology alone cannot replace human-centered instruction (Almashour et al., 2025). Successful implementation depends not merely on access to software, but on the systematic development of critical evaluation skills (Peng, 2026), ethical awareness, and structured AI literacy (Celik, 2023; Celik et al., 2026).

In order to keep AI integration ethically grounded and pedagogically effective within post-method language education, future research should target several key areas. Regarding long-term cognitive and metacognitive effects, longitudinal studies are needed to determine how extended use of GenAI tools

influences lasting language retention, advanced writing skills, and independent problem-solving skills, while also investigating whether continuous automated feedback leads to metacognitive erosion or disrupts self-regulated learning strategies over time.

In terms of teacher identity and agency, future research should investigate how sustained dependence on AI-generated materials shaped EFL teachers' professional authority, independence, and classroom decision-making. Empirical work should also examine structured professional development models, such as lesson study, to identify how pre-service teachers can best build the agency needed to critically adapt AI outputs for diverse contexts.

Addressing the multi-tiered digital divide and algorithmic bias requires future work focused on developing AI platforms specifically designed for under-resourced, multilingual, and indigenous language settings, so as to counteract the Western-centric biases (Bozic, 2023). Researchers also need to examine the third-level digital divide (Matjie et al., 2026), assessing whether unequal quality in AI-generated feedback works to reinforce existing educational disparities across different socioeconomic settings. Finally, achieving the full potential of human-AI collaboration will require additional empirical research to identify the ideal balance between automated real-time feedback and human-mediated dialogic instruction. Researchers should examine how to design AI-supported tasks that lower speaking anxiety while maintaining the emotional engagement, empathy, and social connection provided by human educators.

In the end, the post-method era calls on language teachers to serve as critical facilitators of technology rather than passive users of it. By embedding Generative AI within a human-centered framework guided by local context, ethical awareness, and critical judgement, the field of EFL can make the most of technological innovation while still protecting learner agency and the integrity of sound pedagogy.

Ethical Statement on the Use of Artificial Intelligence

During the preparation of this study, AI-based language models were utilized only for text synthesis, language editing, and improving readability. The AI tools were not used for data creation, collection, or analysis. The authors retain full responsibility for the content, intellectual integrity, and final version of the manuscript.

References

Adewale, O. A., Rane, J., Ogbonna, M. O., & Rane, N. L. (2026b). Pedagogy with generative artificial intelligence: Opportunities and challenges in education. *International Journal of Applied Resilience and Sustainability*, *2*(2), 339–366. https://doi.org/10.70593/deepsci.0202013

Ahmad, M., & Khasawneh, S. (2023). Integrating AI-based virtual conversation partners in enhancing speaking skills in foreign languages: Insights from university students. *Journal of Southwest Jiaotong University*, *58*(5), 575–587. https://doi.org/10.35741/issn.0258-2724.58.5.43

Alm, A. (2024). Exploring Autonomy in the AI Wilderness: Learner Challenges and Choices. *Education Sciences*, *14*(12), 1369. https://doi.org/10.3390/educsci14121369

Almashour, M., Aldamen, H. a. K., & Jarrah, M. (2025). "They know AI, but they also know us": student perceptions of EFL teacher identity in AI-enhanced classrooms in Jordan. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1611147

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Saleh, K. B., Alowais, S. A., Alshaya, O. A., Rahman, I., Yami, M. S. A., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*, *19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Amin, M. Y. M. (2023). AI and Chat GPT in Language Teaching: Enhancing EFL Classroom Support and Transforming Assessment Techniques. *International Journal of Higher Education Pedagogies*, *4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

Aslan, M. (2025). The evolution of language education in the digital age. In S. Kasap (Ed.), *Language in the digital age*(pp. 118–141). Eğitim Yayınevi.

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, *3*(2), 818–824. https://doi.org/10.25082/amler.2023.02.009

Babayiğit, M. V., & Coşkun, Z. (2023). Computer-assisted language learning (CALL) in foreign language education. In S. Kasap (Ed.), *Methodologies and approaches in ELT classrooms* (pp. 43–56). Anı Publishing.

Bacus, R. C. (2021). Teachers' beliefs, praxes, and post-method pedagogy in English Language teaching. *International Journal of Learning Teaching and Educational Research*, *20*(1), 86–102. https://doi.org/10.26803/ijlter.20.1.5

Bahari, A., Han, F., & Strzelecki, A. (2025). Integrating CALL and AIALL for an interactive pedagogical model of language learning. *Education and Information Technologies*, *30*, 14305–14333. https://doi.org/10.1007/s10639-025-13388-w

Banafi, N. (2025). Status of the English language program considering Saudi Vision 2030: from students' perspectives. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1). https://doi.org/10.1186/s40862-024-00311-y

Bhandari, B. L., & Bhandari, L. P. (2025). Exploring teachers' perceptions of post-method pedagogy in ELT. *International Journal of Language and Literary Studies*, *7*(2), 203–216. https://doi.org/10.36892/ijlls.v7i2.2030

Bhar, S. K. (2026). Artificial intelligence in EFL speaking instruction: A systematic review of pedagogical design, affective conditions and instructional input. *Encyclopedia*, *6*(4), 74. https://doi.org/10.3390/encyclopedia6040074

Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology General*, *151*(9), 2250–2258. https://doi.org/10.1037/xge0001181

Bozic, V. (2023). Strengthening EFL learners' autonomy: Exploring the Voscreen app and its impact on encouragement and engagement in technology-based language learning. *Language Education & Technology (LET Journal), 3*(2), 96-109.

Brugliera, P. (2024). The effectiveness of digital learning platforms in enhancing student engagement and academic performance. *Journal of Education Humanities and Social Research*, *1*(1), 26–36. https://doi.org/10.70088/xq3gy756

Bulathwela, S., Pérez-Ortiz, M., Holloway, C., Cukurova, M., & Shawe-Taylor, J. (2024). Artificial intelligence alone will not democratise education: on Educational inequality, techno-solutionism and inclusive tools. *Sustainability*, *16*(2), 781. https://doi.org/10.3390/su16020781

Celik, I., Kontkanen, S., Laru, J., & Dalyanci, A. A. (2025). Co-constructing adaptive lesson plans with GenAI: Pre-service teachers' Intelligent-TPACK and prompt engineering strategies. *Computers & Education*, *241*, 105485. https://doi.org/10.1016/j.compedu.2025.105485

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chen, M. R. A., & Lin, Y. H. (2023). AI chatbot-supported grammar tense learning on EFL students learning effectiveness. *INTED2023 Proceedings* (pp. 1207–1216). IATED.

Chinta, S. V., Wang, Z., Palikhe, A., Zhang, X., Kashif, A., Smith, M. A., Liu, J., & Zhang, W. (2024). AI-Driven Healthcare: A review on ensuring fairness and mitigating bias. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2407.19655

Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education Artificial Intelligence*, *6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197

Conceição, J. M. C., & Van Der Stappen, E. (2025). The impact of AI on inclusivity in higher education: A rapid review. *Education Sciences*, *15*(9), 1255. https://doi.org/10.3390/educsci15091255

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education*, *1*(3), 158–167. https://doi.org/10.47852/bonviewijce42022495

Dai, Y., & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study. *Computer Assisted Language Learning*, *36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

Deep, P. D., Martirosyan, N., Ghosh, N., & Rahaman, M. S. (2025). ChatGPT in ESL higher education: Enhancing writing, engagement, and learning outcomes. *Information*, *16*(4), 316. https://doi.org/10.3390/info16040316

Derakhshan, A. (2025). EFL students' perceptions about the role of generative artificial intelligence (GAI)-mediated instruction in their emotional engagement and goal orientation: A motivational climate theory (MCT) perspective in focus. *Learning and Motivation*, *90*, 102114. https://doi.org/10.1016/j.lmot.2025.102114

Ding, D., & Yusof, A. M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05550-z

Druga, S., Otero, N., & Ko, A. J. (2022). The landscape of teaching resources for AI education. In *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education* (pp. 96–102). https://doi.org/10.1145/3502718.3524782

Du, H., Xing, W., & Pei, B. (2021). Automatic text generation using deep learning: providing large-scale support for online learning communities. *Interactive Learning Environments*, *31*(8), 5021–5036. https://doi.org/10.1080/10494820.2021.1993932

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295x.114.4.864

Farhat, Z. (2026). When AI thinks for us: Unveiling the cognitive and social toll of ChatGPT over-reliance. *Applied Cognitive Psychology*, *40*(3). https://doi.org/10.1002/acp.70205

Ferrara, E. (2023). Fairness and Bias in Artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4615421

Fu, Y., & Hanaki, N. (2026). Do people rely on ChatGPT more than their peers to detect deepfake news? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2608.01540

Gentile, M., Città, G., Perna, S., & Allegra, M. (2023). Do we still need teachers? Navigating the paradigm shift of the teacher's role in the AI era. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1161777

Guo, X. (2025). Facilitator or thinking inhibitor: understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, *33*(5), 3526–3544. https://doi.org/10.1080/10494820.2024.2445177

Han, B., Coghlan, S., Buchanan, G., & McKay, D. (2025). Who is helping whom? Student concerns about AI-teacher collaboration in higher education

classrooms. *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1–32. https://doi.org/10.1145/3711104

Han, X. (2025). Improving pre-service teachers' AI competencies: a quasi experimental study. *Frontiers in Psychology*, *16*, 1642465. https://doi.org/10.3389/fpsyg.2025.1642465

Haque, M. A., & Li, S. (2024). Exploring ChatGPT and its impact on society. *AI And Ethics*, *5*(2), 791–803. https://doi.org/10.1007/s43681-024-00435-4

Hilgendorf, S. K. (2020). History of language-teaching methods. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0510.pub2

Hoang, N. H. (2026). Investigating EFL students' emotional responses in AI-assisted language learning: a Broaden-and-Build Theory approach. *Interactive Learning Environments*, *34*(3), 1046–1061. https://doi.org/10.1080/10494820.2025.2512567

Hoffmann, L. C. (2026). The impact of artificial intelligence-assisted language learning tools on reading comprehension and vocabulary acquisition among secondary school students. *International Journal of Philology, Literary Studies and Language Education*, *1*(1), 18–23.

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

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hwang, W., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts. *Journal of Educational Computing Research*, *61*(5), 1008–1035. https://doi.org/10.1177/07356331221137253

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, *206*, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, *13*, 1049401. https://doi.org/10.3389/fpsyg.2022.1049401

Kamalov, F., Calonge, D. S., & Gurrib, I. (2023). New era of Artificial intelligence in Education: Towards a sustainable Multifaceted Revolution. *Sustainability*, *15*(16), 12451. https://doi.org/10.3390/su151612451

Kannan, J., & Munday, P. (2018). New Trends in Second Language Learning and Teaching through the lens of ICT, Networked Learning, and Artificial

Intelligence. *Círculo De Lingüística Aplicada a La Comunicación*, *76*, 13–30. https://doi.org/10.5209/clac.62495

Karimvand, P. N., Hessamy, G., & Hemmati, F. (2014). The place of postmethod pedagogy in teacher education programs in EFL language centers of Iran. *Iranian Journal of Applied Linguistics (IJAL)*, *17*(2), 59–91.

Kartal, G. (2024). The influence of ChatGPT on thinking skills and creativity of EFL student teachers: a narrative inquiry. *Journal of Education for Teaching International Research and Pedagogy*, *50*(4), 627–642. https://doi.org/10.1080/02607476.2024.2326502

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, M. K. (2024). PBL using AI technology-based learning tools in a Korean ELT university setting. In *AsiaTEFL Proceedings 2023: Papers from the 21st AsiaTEFL Conference* (pp. 133–144). AsiaTEFL.

Kozlova, Y. (2026). Generative AI in Foreign Language Education: A scoping Review of Pedagogical applications and challenges. *Applied Language Learning*, *35*(1), 13–40.

Kumaravadivelu, B. (2003a). A postmethod perspective on English language teaching. *World Englishes*, *22*(4), 539–550. https://doi.org/10.1111/j.1467-971x.2003.00317.x

Kumaravadivelu, B. (2003b). *Beyond methods: Macrostrategies for language teaching*. Yale University Press. https://doi.org/10.2307/j.ctt1np6r2

Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, *35*(4), 537–560. https://doi.org/10.2307/3588427

Kumaravadivelu, B. (2006). Understanding Language Teaching: From method to Postmethod. *Teaching English as a Second or Foreign Language*, *11*(3).

Kundu, A., & Bej, T. (2025). Empowering Students' autonomy in EFL Learning: AI Innovations in Schools of the Global South. *The Electronic Journal of Information Systems in Developing Countries*, *91*(6). https://doi.org/10.1002/isd2.70041

Kusuma, I. P. I., Roni, M., Dewi, K. S., & Mahendrayana, G. (2024). Revealing the potential of ChatGPT for English language teaching: EFL preservice teachers' teaching practicum experience. *Studies in English Language and Education*, *11*(2), 650–670. https://doi.org/10.24815/siele.v11i2.34748

Lee, S., & Jeon, J. (2022). Visualizing a disembodied agent: young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*, *37*(5–6), 1048–1073. https://doi.org/10.1080/09588221.2022.2067182

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A

paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Little, D. (1991). *Learner autonomy: Definitions, issues, and problems*. Authentik.

Liu, T., & Gablasova, D. (2023). Data-driven learning of collocations by Chinese learners of English: a longitudinal perspective. *Computer Assisted Language Learning*, *38*(3), 612–637. https://doi.org/10.1080/09588221.2023.2214605

Liu, G. L., Darvin, R., & Ma, C. (2024). Unpacking the role of motivation and enjoyment in AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation in the Chinese context. *Computers in Human Behavior*, *160*, 108362. https://doi.org/10.1016/j.chb.2024.108362

Luo, X. A. (2026). AI multimodal personalized language-assisted inclusive early childhood education: Improving early reading and classroom interaction ability of SEN children. In *AIETDS 2026: Proceedings of the 2026 International Conference on Artificial Intelligence in Education Technology and Data Science* (pp. 7–11). https://doi.org/10.1145/3810158.3810161

Mac Fadden, I., García-Alonso, E., & Meneses, E. L. (2024e). Science Mapping of AI as an Educational tool Exploring Digital inequalities: A Sociological perspective. *Multimodal Technologies and Interaction*, *8*(12), 106. https://doi.org/10.3390/mti8120106

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00295-9

Marriott, H. R., & Pitardi, V. (2024). One is the loneliest number. . . Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction. *Psychology and Marketing*, *41*(1), 86–101. https://doi.org/10.1002/mar.21899

Martinaj, F. (2024). The Dichotomy Between Oral and Written Assessment Methods from the Perspective of Students Requiring a Second Language for Their Specific Area of Study. *The International Journal of Assessment and Evaluation*, *32*(1), 25–47. https://doi.org/10.18848/2327-7920/cgp/v32i01/25-47

Matjie, M. A., Nethavhani, A., & Matlakala, M. (2026). AI and the digital divide in education. *Frontiers in Computer Science*, *8*. https://doi.org/10.3389/fcomp.2026.1759027

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Moradi, H. (2025). The role of language teachers' perceptions and attitudes in ICT integration in higher education EFL classes in China. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-04524-5

Namaziandost, E., & Rezai, A. (2024). Special issue: Artificial intelligence in open and distributed learning: Does it facilitate or hinder teaching and learning? *The International Review of Research in Open and Distributed Learning*, *25*(3), i–vii. https://doi.org/10.19173/irrodl.v25i3.8070

Nepal, A. (2023). Pedagogical Shift in ELT through Post-method Pedagogy. *Research Journal*, *8*(1), 1–11. https://doi.org/10.3126/rj.v8i1.60985

Noorbehbahani, F., & Oyibo, K. (2026). AI-overdependence and human cognitive decline: Hazards, evidence, and mitigation strategies. *Computers in Human Behavior Reports*, *22*, 101102. https://doi.org/10.1016/j.chbr.2026.101102

Owoseni, A., Kolade, O., & Egbetokun, A. (2024). Applications of generative AI in summative assessment. In *Generative AI in Higher Education* (pp. 97–122). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-60179-8_4

Peng, X. (2026). Generative AI in pre-service EFL/ESL teacher education: A systematic review of empirical studies. *Journal for the Education of Gifted Young Scientists*, *14*(2), 75–89. https://doi.org/10.17478/jegys.1915711

Raimkulova, A., Ybyraimzhanov, K., Halmatov, M., Mailybayeva, G., & Khaimuldanov, Y. (2025). Pre-Service EFL Primary Teachers Adopting GENAI-Powered Game-Based instruction: A Practicum intervention. *Education Sciences*, *15*(10), 1326. https://doi.org/10.3390/educsci15101326

Rees, T. (2022). Non-Human Words: on GPT-3 as a philosophical laboratory. *Daedalus*, *151*(2), 168–182. https://doi.org/10.1162/daed_a_01908

Sajja, R., Sermet, Y., Fodale, B., & Demir, I. (2026). Evaluating AI-powered learning assistants in engineering higher education with implications for student engagement, ethics, and policy. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-39237-5

Salah, M., Abdelfattah, F., Alhalbusi, H., & Mukhaini, M. A. (2024). Me and My AI Bot: Exploring the "AIholic" Phenomenon and University Students' Dependency on Generative AI Chatbots - Is This the New Academic Addiction? *Research Square*. https://doi.org/10.21203/rs.3.rs-3508563/v2

Samsudin, & Allen, C. J. (2026). Integrating generative AI in English education opportunities and challenges for EFL learners. *Journal of Research in English Language Teaching and Linguistics*, *2*(1), 49–66. https://doi.org/10.65431/jrell.v2i1.32

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How Generative AI Influences Students' Self-Regulated Learning and Critical Thinking Skills? A Systematic review. *International Journal of Engineering Pedagogy (iJEP)*, *15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Senen, I. (2024). Effect of method pedagogy and English teaching learning. *Journal of Society Innovation and Development*, *5*(2), 171–190. https://doi.org/10.63924/jsid.v5i2.48

Shadiev, R., Wu, T., & Huang, Y. (2018). Facilitating comprehension of non-native English speakers during lectures in English with STR-texts. *Journal of Computer Assisted Learning*, *34*(1), 94–104. https://doi.org/10.1111/jcal.12217

Shadiev, R., Wu, T., & Huang, Y. (2020). Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. *ReCALL*, *32*(2), 195–212. https://doi.org/10.1017/s0958344020000038

Simos, C., & Katsiampoura, I. (2025). The use of digital tools to enhance inclusion in the classroom. *World Journal of Advanced Engineering Technology and Sciences*, *17*(1), 465–477. https://doi.org/10.30574/wjaets.2025.17.1.1446

Song, Y., Wei, Y., Shen, Y., & Xu, M. (2022). Evaluation of an online oral English teaching model using big data. *Mobile Information Systems*, *2022*, 1–10. https://doi.org/10.1155/2022/7934575

Stockwell, G. (2022). *Mobile assisted language learning: Concepts, contexts and challenges*. Cambridge University Press. https://doi.org/10.1017/9781108652087

Tai, T., & Chen, H. H. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801

Torres, P. J., & Kahveci, Y. E. (2025b). Effectiveness of Artificial Intelligence (AI) in language teaching. *Computers and Education Artificial Intelligence*, *9*, 100522. https://doi.org/10.1016/j.caeai.2025.100522

Vesna, L. (2025). Digital Divide in AI-Powered Education: Challenges and Solutions for Equitable learning. *Journal of Information Systems Engineering & Management*, *10*(21s), 300–308. https://doi.org/10.52783/jisem.v10i21s.3327

Wan, E. W., & Chen, R. P. (2021). Anthropomorphism and object attachment. *Current Opinion in Psychology*, *39*, 88–93. https://doi.org/10.1016/j.copsyc.2020.08.009

Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course. *Library Hi Tech*, *40*(1), 80–97. https://doi.org/10.1108/lht-05-2020-0113

Warschauer, M., & Healey, D. (1998). Computers and language learning: an overview. *Language Teaching*, *31*(2), 57–71. https://doi.org/10.1017/s0261444800012970

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, *29*(18), 24963–24990. https://doi.org/10.1007/s10639-024-12819-4

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChaTGPT for language learning tasks: Experience and Perceptions. *Languages*, *8*(3), 212. https://doi.org/10.3390/languages8030212

Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2023). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*, *4*(2), 300–325. https://doi.org/10.1515/jccall-2022-0032

Yalcinalp, S., Türkoğlu, H., Koc, S. E., & Ersoy, H. (2024). Utilization of Artificial Intelligence in Education: A Perspective on Learning Strategies. In *Artificial intelligence*. https://doi.org/10.5772/intechopen.1005086

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 327–343. https://doi.org/10.1017/s0958344022000039

Yang, L. (2026). Empowering the autonomous learner: How AI-assisted language learning environments shape self-regulation, autonomy, and self-directed behaviors. *Language Teaching Research*. https://doi.org/10.1177/13621688261422129

YarAhmadi, M., & Behbahani, H. K. (2025). The effect of interventionist dynamic assessment on Iranian EFL learners' vocabulary learning and retention: a sociocultural inquiry. *Language Testing in Asia*, *15*(1). https://doi.org/10.1186/s40468-024-00337-6

Yiling, J., Omar, M., & Kamaruzaman, F. M. (2025). Exploring the AI-Enhanced Project-Based Learning for English Language Acquisition: A systematic review of the key elements and emerging technology trends. *International Journal of Learning Teaching and Educational Research*, *24*(2), 636–652. https://doi.org/10.26803/ijlter.24.2.31

Zaman, S., Hussain, M. S., & Tabassam, M. (2024). Use of Artificial intelligence in Education: English language Teachers' identity negotiation in Higher education. *Journal of Asian Development Studies*, *13*(3), 861–869. https://doi.org/10.62345/jads.2024.13.3.70

Zhang, H., Hwang, W., Tseng, S., & Chen, H. S. L. (2019). Collaborative Drama-Based EFL learning in familiar contexts. *Journal of Educational Computing Research*, *57*(3), 697–722. https://doi.org/10.1177/0735633118757731

Zou, B., & Wang, C. (2024). Using an artificial intelligent speaking assessment platform—EAP Talk—to develop EFL speaking skills. In B. Zou & T. Mahy (Eds.), *English for academic purposes in the EMI context in Asia*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-63638-7_12

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, *13*(3). https://doi.org/10.1177/21582440231193818

Savoring Language Through Translanguaging: A Conceptual and Discursive Perspective

Maghsoud Hassani Mehr [1]

Abstract
In applied linguistics, the affective turn has expanded scholarly attention to emotion in language learning, particularly through constructs like foreign language enjoyment and well-being. Yet, the field has paid far less attention to savoring, the agentive process through which individuals attend to, amplify, and prolong positive experiences over time. Drawing on positive psychology and SLA affect research, this conceptual paper argues that translanguaging provides the semiotic architecture through which savoring is discursively enacted. Translanguaging is approached not merely as a pedagogical orientation, but as a meaning and sense-making framework foregrounding lived experience, multimodality, subjectivity, and identity. We introduce a novel model of "savoring-through-translanguaging" centered on a 4-Stage Discursive-Savoring Cycle (Affective Re-Voicing, Present-Moment Amplification, Anticipatory Re-Framing, and Semiotic Entrenchment) accompanied by a micro-analytic qualitative coding matrix. The paper delineates how savoring differs from enjoyment, demonstrates the framework via illustrative micro-discourse vignettes, and identifies systemic boundary conditions (e.g., monolingual language policing, cognitive load). Implications are offered for moving affect research beyond momentary emotion toward tracking savoring as a co-constructed, discursive accomplishment across learning trajectories.
Keywords: savoring, translanguaging, affect in language learning, multilingual meaning-making, positive psychology in SLA

Introduction

Over the past two decades, applied linguistics has witnessed what is often described as an *affective turn*, marked by a growing interest in emotions, well-being, and subjective experience in language learning and teaching (Dörnyei & Ryan, 2015; Pavlenko, 2013). Early work in the field largely emphasized negative emotions such as anxiety, fear, and apprehension, viewing affect primarily as an obstacle to successful language acquisition (Horwitz, 2001; MacIntyre, 2017). More recent scholarship, influenced by positive psychology, has shifted attention toward positive emotions, with constructs such as enjoyment, hope, pride, and well-being becoming central to contemporary discussions of second and foreign language learning (Dewaele et al., 2019; MacIntyre et al., 2019). This shift has

[1] Lecturer, Istanbul University-Cerrahpaşa, School of Foreign Languages, Department of Foreign Languages, English Preparatory Program, ORCID: 0000-0001-9986-6258, *maghsoud.mehr@iuc.edu.tr*, Türkiye

contributed to a more balanced understanding of affect, highlighting how positive emotional experiences can broaden learners' engagement, support motivation, and foster sustained participation in language practices.

Within this expanding body of research, *enjoyment* has emerged as one of the most extensively studied positive emotions in applied linguistics. Research on foreign language enjoyment has demonstrated its close relationship with motivation, willingness to communicate, and active classroom engagement (Dewaele & MacIntyre, 2014; Dewaele et al., 2018). However, the focus on enjoyment has tended to conceptualize positive affect as a relatively immediate, moment-bound emotional response to learning tasks, classroom interaction, or instructional practices. As a result, less attention has been paid to how positive experiences are *processed, extended, revisited, or narrated over time* by learners and teachers. This gap raises an important theoretical question: how do language users actively work with positive emotions beyond the moment of their initial emergence?

This paper argues that the concept of *savoring*, developed in positive psychology, offers a valuable yet underexplored lens for addressing this question in applied linguistics. Savoring refers to the ways individuals attend to, intensify, prolong, and reflect upon positive experiences, whether by anticipating them, dwelling on them in the present, or recalling them afterward (Bryant & Veroff, 2007). Unlike enjoyment, which denotes an emotional state, savoring emphasizes agency, temporality, and meaning-making. Although savoring has begun to attract empirical attention in second language acquisition (e.g., Pan & Zhang, 2021; Wang et al., 2023), existing studies treat it primarily as an internal, individual-level psychological trait measured through self-report scales. Consequently, the field lacks a structural model that accounts for how savoring is discursively enacted, multimodal, and socially co-constructed over time.

At the same time, translanguaging has become a prominent theoretical and pedagogical orientation in applied linguistics, challenging monolingual assumptions about language, learning, and communication (García & Wei, 2014; Li Wei, 2018). Translanguaging foregrounds language users' full semiotic repertoires and treats meaning-making as dynamic, embodied, and socially situated. By emphasizing lived experience, affect, memory, and identity, translanguaging provides a framework for rethinking how positive experiences in language learning are expressed, sustained, and transformed over time (Li Wei, 2022). Yet, despite its attention to subjectivity and experience, translanguaging research has rarely engaged explicitly with savoring as a form of affective–discursive practice.

Against this background, the present paper brings together insights from positive psychology, applied linguistics, and translanguaging theory to propose a conceptual and discursive account of savoring language through translanguaging. Rather than treating savoring as an individual psychological trait, the paper approaches it as an agentive process that is socially and discursively accomplished through multilingual and multimodal resources. By positioning savoring within translanguaging spaces, this conceptual paper seeks to extend current discussions of affect beyond momentary enjoyment, offering new directions for theorizing positive experience, agency, and meaning-making in translanguaging-oriented research.

Statement of the Problem and Purpose of the Paper

Despite the growing prominence of affect in applied linguistics, current theorization remains uneven in how positive experience is conceptualized and examined. While constructs such as enjoyment, motivation, and well-being have been widely investigated, they are most often treated as *emotional states* associated with specific classroom moments or instructional conditions. This emphasis has limited the field's ability to explain how positive experiences are actively worked on, sustained, and transformed over time by learners and teachers. As a result, applied linguistics lacks a robust conceptual vocabulary for addressing the agentive and temporal dimensions of positive affect, how individuals linger on, revisit, narrate, and re-embed positive moments into their ongoing language learning trajectories.

Savoring, a construct well established in positive psychology, directly addresses this gap by focusing on how people attend to, prolong, and make meaning out of positive experiences. However, savoring has not yet been fully integrated into applied linguistics theories, and where it has appeared, it has often been approached through individual-level or psychological lenses. This has limited its potential to illuminate the social, discursive, and semiotic processes through which positive experiences are enacted and shared in language education settings. In particular, there remains a lack of conceptual work that situates savoring within broader sociocultural frameworks that foreground language, interaction, identity, and power.

At the same time, translanguaging research has significantly reshaped understandings of language, learning, and meaning-making by emphasizing learners' full semiotic repertoires and lived experiences. Translanguaging perspectives explicitly foreground affect, subjectivity, memory, and agency, yet they have rarely been mobilized to theorize how positive

experiences are sustained and reworked over time. This disconnect represents a missed opportunity to link translanguaging's emphasis on lived, multimodal meaning-making with a richer account of positive affective processes such as savoring.

The purpose of this paper is to move beyond descriptive broad outlines by establishing a granular, semiotically grounded framework for savoring-through-translanguaging. Specifically, the paper aims to (a) establish a comparative matrix differentiating Foreign Language Enjoyment, psychological savoring, and savoring-through-translanguaging, (b) detail a 4-Stage Discursive-Savoring Coding Scheme with micro-analytic indicators for SLA researchers, (c) evaluate theoretical vignettes demonstrating the framework in action, and (d) outline the systemic boundary conditions where translanguaging fails to facilitate affective savoring. By doing so, the paper seeks to extend current affect-oriented research beyond momentary emotion and to contribute a theoretically grounded perspective on how learners and teachers come to *savor language* through translanguaging practices.

Literature Review

The Affective Turn in Applied Linguistics

The shift from deficit-oriented perspectives on foreign language anxiety toward positive psychology in applied linguistics has established positive emotions—such as enjoyment, hope, and resilience—as central to learning and engagement (MacIntyre et al., 2019). However, this literature remains heavily anchored in momentary emotional states measured during or immediately after instruction (Dewaele & MacIntyre, 2014). Consequently, the field lacks theoretical mechanisms to explain how language users actively extend, revisit, and discursively transform positive affective experiences over time.

Building on this broader affective orientation, the field has more recently been shaped by the positive psychology turn in second language acquisition and applied linguistics. Drawing explicitly on developments in psychology, this line of research foregrounds positive emotional experiences and learner strengths, including enjoyment, hope, resilience, grit, and well-being (MacIntyre et al., 2019). Within this framework, positive emotions are understood not merely as pleasant by-products of learning but as resources that can broaden learners' attention, enhance engagement, and support sustained motivation over time. This perspective has significantly rebalanced affective research, which had previously been dominated by negative constructs.

Among the various positive affective constructs introduced during this period, foreign language enjoyment has received the most sustained empirical attention. Enjoyment has been conceptualized as a complex, socially mediated emotion arising from meaningful challenge, supportive classroom relationships, and a sense of accomplishment in language use (Dewaele & MacIntyre, 2014). A growing body of research has demonstrated robust associations between enjoyment and motivation, willingness to communicate, classroom engagement, and perceived learning success (Dewaele et al., 2018; Dewaele et al., 2019). Despite these contributions, enjoyment has typically been examined as an *immediate emotional state* tied to particular tasks, interactions, or instructional moments.

This emphasis on momentary emotional experience reveals a conceptual limitation within the affective turn in applied linguistics. While the field has become increasingly attentive to *what learners feel*, it has paid comparatively little attention to *what learners do with those feelings over time*. Many affective constructs adopted from psychology, including enjoyment, motivation, and well-being, retain a predominantly individual and state-oriented focus. Because current models treat positive affect as an episodic outcome rather than an ongoing process, they fail to account for the semiotic and discursive mechanisms through which learners actively linger on and re-entextualize moments of success across their learning trajectories.

Savoring directly addresses this gap. Originating in positive psychology, savoring refers to the agentive processes through which individuals attend to, intensify, prolong, and reflect on positive experiences (Bryant & Veroff, 2007). Psychological research on savoring has examined strategies such as anticipation, present-moment awareness, and reminiscence, emphasizing how individuals regulate and extend positive affect. Although some scholars have begun to introduce savoring into educational and second language learning contexts, it has largely remained conceptualized as an individual psychological capacity or belief system (Pan & Zhang, 2021). As a result, savoring has not yet been fully integrated into applied linguistics' broader concerns with discourse, interaction, identity, and social meaning-making.

In sum, the affective turn in applied linguistics has progressed from a narrow focus on negative emotions to a richer engagement with positive affect and well-being. However, this progression has largely remained anchored in psychological constructs that prioritize emotional states rather than affective processes. While enjoyment has become a central concept, savoring, despite its relevance to learning, agency, and temporality, remains marginal in applied linguistics theorization. This conceptual gap

points to the need for frameworks that can account for how positive experiences are socially mediated, discursively enacted, and sustained over time. In the next section, savoring is examined in greater detail as a construct rooted in psychology but increasingly relevant to language learning and teaching

Conceptualizing Savoring in Positive Psychology and Language Learning

The concept of savoring originates firmly in positive psychology, where it was introduced to explain how individuals actively engage with positive experiences rather than merely experiencing them passively. Seminal work by Bryant (2003) and Bryant and Veroff (2007) conceptualized savoring as the processes through which people attend to, intensify, and prolong positive emotional experiences across time before, during, and after an event. In this framework, savoring is not itself an emotion but a meta-affective process that regulates emotional experience through anticipation, present-moment awareness, and reminiscence (Bryant & Veroff, 2007).

Within psychology, savoring has been widely examined in relation to well-being, resilience, and life satisfaction. Empirical studies have demonstrated that individuals who frequently engage in savoring strategies report higher levels of sustained positive affect and psychological well-being (Bryant et al., 2011). Importantly, this line of research clearly distinguishes savoring from related constructs such as pleasure or enjoyment by emphasizing *process rather than emotional state*. Whereas enjoyment refers to the experience of positive emotion, savoring refers to how individuals *work on* that emotion to extend its duration and meaning (Bryant, 2003).

More recently, savoring has begun to appear in educational and learning-related research, including studies situated within second language learning and teaching. Drawing largely on positive psychology frameworks, these studies have explored savoring in terms of learners' beliefs, dispositions, and strategies for engaging with positive learning experiences (Pan & Zhang, 2021). Findings suggest that savoring can support motivation, emotional regulation, persistence, and learner well-being, positioning it as a potentially valuable construct within the broader positive psychology movement in second language acquisition (MacIntyre et al., 2019).

Despite these promising developments, the introduction of savoring into language learning research has so far remained conceptually narrow. Most existing studies conceptualize savoring as an individual psychological

mechanism, focusing on internal cognitive or emotional processes rather than on interaction, discourse, or social context (Pan & Zhang, 2021). As a result, savoring is often treated as detached from the communicative practices through which learning experiences are expressed, shared, and remembered. This limitation reflects a broader tendency within applied linguistics to import affective constructs from psychology without fully re-theorizing them in relation to language, social interaction, and meaning-making (Pavlenko, 2013; Prior, 2019).

From an applied linguistics perspective, this reliance on psychologically oriented conceptualizations leaves important questions unanswered. In particular, it remains unclear how positive experiences are discursively constructed, narrated, and recontextualized through language use over time. Scholars working within sociocultural, discourse-analytic, and ecological approaches have repeatedly emphasized that affect is not merely an internal state but is co-constructed through interaction, positioning, and semiotic resources (Larsen-Freeman, 2018; Prior, 2019). Without attending to these dimensions, savoring risks remaining an imported psychological construct rather than becoming an analytically productive concept within applied linguistics.

In summary, savoring is well established in psychology as an agentive process of engaging with positive experience (Bryant & Veroff, 2007; Bryant et al., 2011), and it has begun to attract attention in language learning research through positive psychology lenses (Pan & Zhang, 2021; Wang et al., 2023). However, applied linguistics still lacks a robust conceptualization of savoring that foregrounds discourse, interaction, and social meaning-making. Addressing this gap requires a framework that can situate savoring within lived multilingual experience and semiotic practice. The following section argues that *translanguaging* provides such a framework by enabling an affective–discursive understanding of how positive language learning experiences are sustained and transformed.

Translanguaging as an Affective–Discursive Framework

Moving beyond code-based models, translanguaging conceptualizes meaning-making as a dynamic, embodied process where speakers deploy their full semiotic repertoires to enact social and affective stance (Garcia & Wei, 2014; Li Wei, 2018). From this perspective, translanguaging is not merely about alternating between named languages but about engaging diverse linguistic, embodied, and multimodal resources in situated social action.

Importantly, translanguaging scholarship has increasingly foregrounded lived experience and subjectivity, positioning affect as integral to

meaning-making rather than as a peripheral variable. García and Wei (2014) emphasize that translanguaging practices are embedded in speakers' histories, identities, and emotional investments, while Li Wei (2018) argues that translanguaging challenges structuralist views of language by centering human experience and creativity. This shift aligns translanguaging closely with broader sociocultural and ecological approaches in applied linguistics that view language learning as an emergent, relational, and affect-laden process (Larsen-Freeman, 2018).

More recently, translanguaging has been advanced not only as a theory or pedagogy but also as a methodological and epistemological framework. Li Wei (2022) proposes "translanguaging as method," arguing that translanguaging directs analytical attention to sense-making, lived experience, memory, feeling, and subjectivity. From this standpoint, affect is not treated as an internal psychological state to be measured but as something enacted through semiotic orchestration in real time. This methodological orientation invites researchers to ask different questions about how emotions are experienced, expressed, and sustained through language use across contexts.

Despite this growing attention to affective dimensions, translanguaging research has tended to conceptualize emotions implicitly rather than explicitly theorizing them. Positive emotional outcomes are often reported as benefits or by-products of translanguaging pedagogy, such as enjoyment or reduced anxiety, without detailed examination of how these experiences are discursively constructed, extended, or reflected upon over time (Duarte, 2020; Yüzlü & Dikilitaş, 2022). Consequently, trans-languaging scholarship has yet to fully engage with processes like savoring, which foreground the agentive and temporal work individuals do with positive experience.

From an affective–discursive perspective, translanguaging offers a particularly productive framework for reconceptualizing savoring. Translanguaging spaces enable learners and teachers to narrate experiences, revoice achievements, revisit moments of success, and layer affective meaning across languages and modes. These practices align closely with psychological accounts of savoring as involving anticipation, present-moment attention, and reminiscence (Bryant & Veroff, 2007), while simultaneously situating savoring within social interaction and discourse. In this sense, translanguaging does not simply facilitate positive emotion; it provides the semiotic resources through which positive experiences can be sustained, reworked, and shared.

Furthermore, translanguaging's emphasis on agency and repertoire challenges individualistic models of affect by highlighting how savoring can

be co-constructed. Learners may savor positive experiences not only privately but also collectively, through shared narratives, peer interaction, and teacher–student dialogue that spans multiple languages and semiotic forms (García & Wei, 2014; Li Wei, 2022). This perspective resonates with applied linguistics research that conceptualizes agency as distributed and relational rather than solely individual (Mercer, 2012; Toohey, 2019).

In summary, translanguaging provides a theoretically rich and methodologically flexible framework for understanding affect as discursively and socially accomplished. While translanguaging research has acknowledged positive affective outcomes, it has not yet explicitly theorized how learners and teachers engage in *savoring* as an agentive process. By bringing savoring into dialogue with translanguaging, applied linguistics can move beyond momentary accounts of enjoyment toward a deeper understanding of how positive language learning experiences are sustained over time through multilingual and multimodal meaning-making. This integration lays the groundwork for the conceptual framework developed in the next section.

Conceptual Framework: Savoring Language Through Translanguaging

Building on insights from positive psychology, applied linguistics, and translanguaging theory, this paper proposes a conceptual framework that understands savoring language through translanguaging as an affective–discursive process. In contrast to approaches that treat positive emotion as an internal, momentary state, this framework conceptualizes savoring as a socially situated activity through which learners and teachers actively attend to, prolong, and rework positive language learning experiences using multilingual and multimodal resources (Bryant & Veroff, 2007; Li Wei, 2022).

At the core of this framework is the distinction between *enjoyment* and *savoring*. Enjoyment refers to the immediate experience of positive emotion during language learning activities, such as pleasure derived from successful communication or engaging classroom interaction (Dewaele & MacIntyre, 2014). Savoring, by contrast, involves the *agentive work* that follows or surrounds such experiences, anticipating success, dwelling on achievement, recalling meaningful moments, and narrating them across time (Bryant, 2003). This temporal and process-oriented nature of savoring aligns with applied linguistics perspectives that view learning as dynamic and emergent rather than episodic (Larsen-Freeman, 2018).

Translanguaging provides the semiotic and discursive conditions through which savoring can be enacted. By legitimizing the use of learners' full linguistic repertoires, translanguaging expands the range of resources available for expressing affect, reflecting on experience, and constructing meaning (Garcia & Wei, 2014). Learners may, for example, recount a successful interaction in a home language, elaborate feelings of pride through mixed-language narratives, or revisit classroom achievements through multimodal artifacts such as notes, recordings, or peer dialogue. These practices allow positive experiences to be revisited and re-entextualized, key features of savoring as described in psychological research (Bryant et al., 2011).

To delineate how savoring-through-translanguaging diverges from previous affect constructs in applied linguistics, Table 1 outlines the conceptual boundaries separating Foreign Language Enjoyment (FLE), psychological savoring, and the proposed affective–discursive framework.

Table 1: *Conceptual Matrix of Affective Constructs in SLA*

Dimension	Foreign Language Enjoyment (FLE)	Standalone Savoring (Psychological)	Savoring-Through-Translanguaging (Affective–Discursive)
Primary Nature	Momentary affective state	Meta-affective cognitive process	Socially & discursively co-constructed practice
Linguistic Scope	Often bound to target-language (L2) performance	Internal, non-linguistic or L1-bound	Full semiotic repertoire (L1 + L2 + multimodal)
Temporal Focus	Immediate present (in situ)	Past, present, or future (internal)	Dynamic re-entextualization across learning trajectories
Primary Agency	Passive/reactive to task or environment	Individual psychological regulation	Relational, distributed, and ideologically mediated

The 4-Stage Discursive-Savoring Cycle

Savoring-through-translanguaging operates through a four-stage cyclic mechanism through which positive affective experiences are transformed into sustained learning investments:

1. Affective Re-Voicing (Retrospective Savoring): This stage refers to the agentive process wherein learners revisit past L2 accomplishments or overcome linguistic anxiety by re-narrating the event through their home language (L1), regional idioms, or affective markers. Drawing on psychological models of reminiscence (Bryant & Veroff, 2007) and applied linguistics perspectives on emotional voice (Pavlenko, 2005, 2013), this stage accounts for how switching to L1 allows learners to re-voice and savor the depth of their achievement when L2 emotional vocabulary is insufficient.

2. Present-Moment Amplification (Concurrent Savoring): This stage involves lingering on a successful L2 moment right as it occurs in interaction. Grounded in Bryant and Veroff's (2007) concept of "marveling" and SLA work on interactional dynamics (Dewaele & MacIntyre, 2014), learners and peers deploy dual-language repetition (echoing an L2 term alongside an L1 equivalent), shared laughter, and paralanguage (tone shifts, sighing, high-fives) to collectively prolong positive affect.

3. Anticipatory Re-Framing (Prospective Savoring): This stage occurs prior to high-stakes or anxiety-inducing L2 tasks. Synthesizing prospective savoring (Bryant, 2003) with translanguaging as a psychological safety net (Garcia & Wei, 2014; Creese & Blackledge, 2010), learners utilize translanguaging spaces to pre-experience positive emotions and reduce threat by mixing L1 collaborative planning with L2 practice lines. Through cross-linguistic humor and peer scaffolding, learners transform anticipatory anxiety into collaborative confidence.

4. Semiotic Entrenchment (Identity Consolidation): The final stage converts momentary savoring into long-term multilingual identity. Drawing on entextualization theory (Prior, 2019) and translanguaging approaches to identity construction (Li Wei, 2018, 2022), temporary interactional moments are locked into permanent physical or digital artifacts (dual-language reflection journals, audio logs, digital portfolios). Over time, this shifts the learner's self-concept from a struggling L2 speaker to a competent multilingual user.

Micro-Analytic Coding Framework for Savoring-Through-Translanguaging

To operationalize savoring-through-translanguaging for empirical analysis, Table 2 outlines observable discourse markers and qualitative coding tags ([BRACKETED]). To assist qualitative researchers, three key analytical concepts underpinning this framework are defined as follows:

- Paralanguage ([PARALANG]): Non-verbal or vocal vocalizations that accompany speech, including pitch variations, laughter, sighs of relief, and prolonged pauses.
- Dual-Language Repetition ([DUAL-REP]): The immediate re-saying of an L2 word or phrase in the L1 (or vice versa) to emphasize meaning and celebrate accuracy.
- Entextualization ([ARTIFACT]): Rendering spoken, temporary interaction into permanent written or recorded form (e.g., dual-language entries).

Table 2: *Micro-Analytic Coding Matrix for Savoring-Through-Translanguaging*

Stage	What the Learner Does (Primary Function)	Observable Classroom & Discourse Markers	Sample Qualitative Research Codes
1. Affective Re-Voicing	Retells a past L2 success or overcoming difficulty with emotion	• Switching to home language (L1) to express emotional relief	• [REVOICE-L1-EMOT] *(L1 Emotional Expression)*
		• Re-narrating past task moments using L1/L2 mix	• [REVOICE-RETRO] *(Retrospective Storytelling)*
		• Softened tone of voice, laughter, or relieved sighing	• [REVOICE-PARALANG] *(Sighs/Tone Shifts)*
2. Present-Moment Amplification	Celebrates and lingers on a successful L2 moment right as it happens	• Repeating the successful word/phrase across L1 and L2	• [AMP-DUAL-REP] *(Dual-Language Repetition)*
		• Shared peer laughter and overlapping celebration talk	• [AMP-LAUGH-OVERLAP] *(Shared Peer Laughter)*
		• Physical gestures (high-fives, chest-taps, smiles)	• [AMP-GESTURE] *(Multimodal Gestures)*
3. Anticipatory Re-Framing	Uses home language to reduce anxiety before a scary L2 task	• Planning in L1 while practicing lines in L2	• [ANTIC-L1-SCAF] *(L1 Planning Scaffolding)*

		• Using cross-linguistic jokes/humor to diffuse stress • Peer-to-peer encouragement and verbal safety building	• [ANTIC-HUMOR] *(Stress-Relieving Humor)* • [ANTIC-AFF-SAFETY] *(Safety Building)*
4. Semiotic Entrenchment	Locks the success into memory to build long-term multilingual confidence	• Saying statements like: "I feel confident in both languages" • Writing dual-language journal entries or recording reflection audio • Long-term positive shift in how the student views themselves	• [ENTRENCH-POS] *(Self-Positioning Statement)* • [ENTRENCH-ARTIFACT] *(Dual-Language Journal/Audio)* • [ENTRENCH-ID-SHIFT] *(Identity Shift)*

Operationalizing the Framework: Illustrative Vignettes

To demonstrate how the 4-Stage Discursive-Savoring Cycle and micro-analytic coding matrix operate in practice, this section presents two theoretical vignettes modeled on typical interactions in multilingual language learning contexts.

Vignette 1: Present-Moment Amplification & Affective Re-Voicing

Setting: *Two intermediate language learners (Seda and Arda) working on a paired oral task.*

Seda (L2): *"Our research proves that environment is... sustainable."*

Arda (L1 + L2): *"Woah, sustainable! Demin bu kelimeyi hatırlayamıyordun, harika söyledin!" (Woah, sustainable! You couldn't remember this word a minute ago; you said it amazingly!)* [AMP-DUAL-REP] [AMP-LAUGH-OVERLAP]

Seda (L1 + Paralanguage): *(Laughs, taps chest) "Valla içim rahatladı, 'sustainability' kelimesinden nefret ediyordum ama sonunda başardım!" (Honestly, I feel so relieved; I used to hate the word 'sustainability' but I finally nailed it!)* [REVOICE-L1-EMOT] [REVOICE-PARALANG]

Arda (L2 + L1): *"Yes! You nailed it valla!"* [AMP-DUAL-REP]

Micro-Discourse Analysis: In Line 1, Seda produces a complex target-language item (sustainable). In Line 2, rather than moving immediately to task completion, Arda initiates Present-Moment Amplification ([AMP-DUAL-REP]) by interrupting to validate the breakthrough. In Line 3, Seda engages in Affective Re-Voicing ([REVOICE-L1-EMOT]); she switches to her L1 to express emotional relief ("içim rahatladı") and reflect on past frustration. The translanguaging space allows both learners to linger on and prolong positive affect before returning to task demands.

Vignette 2: Anticipatory Re-Framing & Semiotic Entrenchment

Setting: *A multilingual learner (Elena) preparing for an upcoming group presentation and writing a post-task reflection.*

Elena (Pre-task peer exchange - L1): *"Şimdi sunuma başlarken heyecanlanırsam Türkçe bir espri yapıp toparlayacağım, tamam mı?" (If I get nervous starting the presentation, I'll make a quick joke in Turkish to regain control, okay?)* [ANTIC-HUMOR] [ANTIC-AFF-SAFETY]

Elena (Post-task dual-language reflection entry): *"Today I spoke for 5 minutes in English without freezing.* [L1 insert]: *Eskiden heyecanlanınca tıkanıyordum ama artık iki dili de kullanabildiğimi bilmek bana güven veriyor. I am becoming a real bilingual speaker."* [ENTRENCH-POS] [ENTRENCH-ARTIFACT] [ENTRENCH-ID-SHIFT]

Micro-Discourse Analysis: In Line 1, Elena uses Anticipatory Re-Framing ([ANTIC-HUMOR]) to reduce pre-performance anxiety by establishing a safety net. In Line 2, she records her success in a dual-language reflection log (Semiotic Entrenchment). By entextualizing the moment into a semiotic artifact ([ENTRENCH-ARTIFACT]), she consolidates a long-term identity shift ([ENTRENCH-ID-SHIFT]) from an anxious student to a confident multilingual user.

Boundary Conditions and Systemic Tensions

While the proposed framework highlights the affective potential of translanguaging, savoring-through-translanguaging does not occur automatically in all language learning environments. Its enactment is constrained by systemic, ideological, and cognitive boundary conditions:

1. Institutional Language Policing and Monolingual Ideologies**:** In classrooms strictly governed by "English-Only" policies, translanguaging practices are often penalized or stigmatized as linguistic deficits. When learners face institutional surveillance or shame for using their home languages, attempting to translanguage induces anxiety rather than psychological safety, effectively preventing affective savoring.

2. Peer Positioning and Social Dynamics: Savoring is a relationally mediated process. If peers respond to a learner's translanguaging moves with judgment, mockery, or exclusion, the interactional space collapses. Present-moment amplification relies on shared peer validation; without positive alignment from interlocutors, positive affect cannot be prolonged.

3. Cognitive Overload in High-Stakes Environments: In situations requiring intense, rapid target-language processing (such as real-time standardized speaking exams), switching between semiotic systems can increase cognitive load. When processing capacity is consumed entirely by task demands, learners lack the attentional resources required to linger on or savor success.

Discussion

The conceptual framework proposed in this paper advances current work on affect in applied linguistics by shifting attention from momentary emotional states to the specific discursive processes through which positive experiences are sustained and made meaningful over time (Bryant & Veroff, 2007; Li Wei, 2022). While the affective and positive psychology turns have productively foregrounded constructs like Foreign Language Enjoyment (FLE) and well-being, these have largely been conceptualized as individual, moment-bound psychological reactions (Dewaele & MacIntyre, 2014; Dewaele et al., 2018; MacIntyre et al., 2019). By operationalizing savoring as an affective–discursive process through translanguaging, this paper extends positive psychology in SLA toward a dynamic, agentive, and interactionally grounded model (Larsen-Freeman, 2018; Prior, 2019).

First, the 4-Stage Discursive-Savoring Cycle (Affective Re-Voicing, Present-Moment Amplification, Anticipatory Re-Framing, and Semiotic Entrenchment) solves a key theoretical limitation in current SLA literature: explaining how positive affect moves across time. Rather than treating positive emotion as a passive by-product of task success, the cycle demonstrates how learners deploy their full semiotic repertoires to actively linger on, prolong, and consolidate positive moments (Bryant et al., 2011; García & Wei, 2014). Translanguaging serves as the essential linguistic mechanism that bridges immediate, high-arousal moments of success with long-term identity shift (Li Wei, 2018, 2022).

Second, as demonstrated in the micro-discourse vignettes, this framework shifts the locus of agency from an internal, isolated psychological trait (Bryant, 2003; Pan & Zhang, 2021) to a relational and co-constructed interactional achievement (Mercer, 2012; Toohey, 2019). In Vignette 1,

Seda's ability to savor her target-language breakthrough relies directly on Arda's peer amplification ([AMP-DUAL-REP]) and the shared translanguaging space that legitimizes her emotional re-voicing ([RE-VOICE-L1-EMOT]). Savoring is thus revealed to be socially mediated; it depends on communicative alignment between interlocutors rather than isolated cognitive self-regulation (Duarte, 2020; Yüzlü & Dikilitaş, 2022).

Finally, the explicit identification of boundary conditions and systemic tensions prevents an overly idealized view of translanguaging. Translanguaging does not automatically generate positive affect. Where institutional language policing enforces strict monolingual norms, or where peer dynamics involve mockery, the interactional space collapses, turning potential savoring moments into sources of anxiety or surveillance (Horwitz, 2001; Pavlenko, 2013; Toohey, 2019). Similarly, under high cognitive load (such as high-stakes oral exams), the processing demands of code-switching can impede the attentional focus required to linger on success (MacIntyre & Gardner, 1994). Savoring-through-translanguaging must therefore be understood as an ideologically and contextually conditional practice.

Implications for Applied Linguistics and Language Education

The operationalization of savoring-through-translanguaging yields distinct implications for theoretical development, empirical research methodology, and pedagogical practice:

Theoretical and Methodological Implications

• From State Measurement to Micro-Analytic Discourse Analysis: Researchers studying SLA affect can move beyond self-report survey scales (e.g., measuring FLE or savoring beliefs via Likert items; Bryant, 2003; Pan & Zhang, 2021). Table 2 Micro-Analytic Coding Matrix offers qualitative researchers concrete, observable indicators, such as dual-language repetition ([AMP-DUAL-REP]), paralanguage ([REVOICE-PARA-LANG]), and artifact entextualization ([ENTRENCH-ARTIFACT]), to analyze classroom transcripts, audio logs, and video recordings (Prior, 2019).

• Operationalizing "Translanguaging as Method": This framework provides an empirical model for operationalizing Li Wei's (2022) concept of translanguaging as a methodological lens, directly linking semiotic resource deployment to emotional regulation and identity construction over time (García & Wei, 2014; Li Wei, 2018).

Pedagogical Implications

• Designing for "Savoring Windows": Rather than rushing learners immediately from one task to the next upon correct completion, teachers can intentionally build in brief "savoring windows" (Bryant & Veroff, 2007). Instructors can encourage peer validation, dual-language echoing, and brief L1 reflections to allow students to linger on breakthroughs (Duarte, 2020).

• Scaffolding Anticipatory Safety: Teachers can utilize Anticipatory Re-Framing ([ANTIC-HUMOR], [ANTIC-L1-SCAF]) before anxiety-inducing tasks by permitting home-language planning and humorous rehearsal to lower affective filters (Creese & Blackledge, 2010; MacIntyre et al., 2019).

• Entextualizing Multilingual Artifacts: To facilitate Semiotic Entrenchment, educators can implement dual-language reflection logs, audio reflection entries, or digital portfolios (Li Wei, 2022; Prior, 2019). Transforming temporary spoken achievements into permanent written or recorded artifacts helps learners consolidate long-term multilingual self-concepts.

• Dismantling Institutional Language Policing: Creating environments where savoring can occur requires actively dismantling rigid monolingual ideologies (Li Wei, 2018; Toohey, 2019). When schools stigmatize home-language use, they strip learners of vital semiotic tools needed to express relief, take pride in learning, and build psychological resilience (Pavlenko, 2013).

Conclusion

This conceptual paper set out to address a critical theoretical gap within the affective turn in applied linguistics by introducing savoring as a vital, missing construct and demonstrating how translanguaging provides the semiotic architecture for its discursive enactment (Bryant & Veroff, 2007; Li Wei, 2018, 2022). While foreign language positive psychology has productively expanded scholarly focus toward emotions such as enjoyment, hope, and well-being, these have predominantly been conceptualized as momentary emotional reactions measured in situ (Dewaele & MacIntyre, 2014; Dewaele et al., 2018; MacIntyre et al., 2019). By contrast, savoring foregrounds the agentive, temporal, and meaning-making work through which language users actively attend to, intensify, prolong, and consolidate positive learning experiences over time (Bryant, 2003; Bryant et al., 2011).

Synthesizing positive psychology, applied linguistics, and translanguaging theory, the paper conceptualized savoring not as an internal cognitive trait (Pan & Zhang, 2021) but as a socially mediated, affective–discursive accomplishment (Larsen-Freeman, 2018; Prior, 2019). Through the introduction of the 4-Stage Discursive-Savoring Cycle (Affective Re-Voicing, Present-Moment Amplification, Anticipatory Re-Framing, and Semiotic Entrenchment), this paper has provided a granular matrix detailing how multilingual resources, paralanguage, dual-language repetition, and multimodal entextualization operate to turn fleeting task success into sustained multilingual identity shifts (Li Wei, 2018; Prior, 2019). Furthermore, by analyzing concrete theoretical vignettes and contextual boundary conditions, the paper highlighted that savoring-through-translanguaging is relationally co-constructed (Mercer, 2012; Toohey, 2019) and constrained by institutional language policing and cognitive load (Horwitz, 2001; Pavlenko, 2013).

Ultimately, this framework contributes to applied linguistics by offering a theoretical and analytical bridge connecting emotion, agency, discourse, and temporality. Methodologically, the proposed micro-analytic matrix opens new avenues for qualitative researchers to operationalize "translanguaging as method" (Li Wei, 2022) by tracing real-time classroom interactions, reflection logs, and audio artifacts. Pedagogically, it urges educators to cultivate deliberate "savoring windows" and dismantle rigid monolingual ideologies that strip learners of their home-language emotional resources (Creese & Blackledge, 2010; Duarte, 2020). By moving beyond how learners feel to examine how they carry those feelings forward through their full semiotic repertoires, applied linguistics can better understand how multilingual learners come to savor language across their educational trajectories.

AI Use Disclosure

During the preparation of this manuscript, the author used ChatGPT (OpenAI) as a language-support tool to improve the clarity, academic style, and grammatical accuracy of the text. All the content of the manuscript, including the conceptualization of the study, literature selection and interpretation, analysis, and scholarly arguments were developed solely by the author.

References

Bryant, F. B. (2003). Savoring beliefs inventory (SBI): A scale for measuring beliefs about savoring. *Journal of Mental Health, 12*(2), 175–196. https://doi.org/10.1080/0963823031000103489

Bryant, F. B., Chadwick, E. D., & Kluwe, K. (2011). Understanding the processes that regulate positive emotional experience: Unsolved problems and future directions for theory and research on savoring. *Frontiers in Psychology, 2*, Article 49. https://doi.org/10.3389/fpsyg.2011.00049

Bryant, F. B., & Veroff, J. (2007). *Savoring: A new model of positive experience*. Lawrence Erlbaum Associates.

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal, 94*(1), 103–115. https://doi.org/10.1111/j.1540-4781.2009.00986.x

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research, 22*(6), 676–697. https://doi.org/10.1177/1362168817692161

Dewaele, J.-M., Chen, X., Padilla, A. M., & Lake, J. (2019). The flowering of positive psychology in foreign language teaching and acquisition research. *Frontiers in Psychology, 10*, Article 2128. https://doi.org/10.3389/fpsyg.2019.02128

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Duarte, J. (2020). Translanguaging in the context of mainstream multilingual education. *International Journal of Multilingualism, 17*(2), 232–247. https://doi.org/10.1080/14790718.2018.1512607

Garcia, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics, 21*, 112–126. https://doi.org/10.1017/S0267190501000071

Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals, 51*(1), 55–72. https://doi.org/10.1111/flan.12314

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amw054

Li, W. (2022). Translanguaging as method. *Applied Linguistics, 43*(1), 1–25. https://doi.org/10.1093/applin/amab058

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

MacIntyre, P. D., Gregersen, T., & Mercer, S. (2019). *Positive psychology in SLA*. Multilingual Matters.

Mercer, S. (2012). The complexity of learner agency. *Applied Linguistics, 33*(1), 41–59. https://doi.org/10.1093/applin/amr039

Pan, Y., & Zhang, L. J. (2021). Learner beliefs about savoring and their role in second language learning. *System, 100*, Article 102548. https://doi.org/10.1016/j.system.2021.102548

Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge University Press.

Pavlenko, A. (2013). The affective turn in SLA: From "affective factors" to "language desire" and "commodification of affect." In D. Gabryś-Barker & J. Bielska (Eds.), *The affective dimension in second language acquisition* (pp. 3–28). Multilingual Matters.

Prior, P. (2019). *Writing/disciplinarity: A sociohistoric account of literate activity in the academy*. Routledge.

Toohey, K. (2019). Disrupting potentialities: Working with the complexities of agency in education. *Language and Education, 33*(5), 459–469. https://doi.org/10.1080/09500782.2019.1623177

Wang, Y., Derakhshan, A., & Zhang, L. J. (2023). Researching and practicing positive psychology in second/foreign language learning and teaching: The past, current status, and future directions. Frontiers in Psychology, 14, Article 1161272. https://doi.org/10.3389/fpsyg.2023.1161272

Yüzlü, M. Y., & Dikilitaş, K. (2022). Translanguaging in the development of EFL learners' foreign language skills in Turkish context. *Innovation in Language Learning and Teaching, 16*(2), 176–190. https://doi.org/10.1080/17501229.2021.1892698

Part II
The Teacher

Learning to See Before Learning to Teach: The Unnamed Role in the Practicum of Pre-Service Teachers of German in Türkiye

Ebru Üstün [1] & Bilal Üstün [2]

Abstract
Pre-service teachers of German in Türkiye spend a substantial part of their final year of study in classrooms, and only a small part of this time is devoted to their own teaching. The national regulation of 2021 specifies the hours in which pre-service teachers teach, while the remaining hours have neither an assigned task nor a designation in the regulatory text, even though the regulation requires them to be planned. These hours are the focus of this chapter. Based on the national regulation, four institutional practicum documents from Turkish universities, and research on professional vision, the chapter shows that classroom presence produces professional learning only where observation is structured. Noticing emerges as knowledge-dependent rather than as a generic ability to look closely; the question of subject-specific observation has been examined in other subjects but remains largely open for foreign language teaching; and neither the existing observation instruments nor the established *clinical supervision model* addresses the hours without the pre-service teachers' own teaching. On this basis, the chapter proposes a framework that organizes observation along three axes, the object of perception, the form of mediation, and the epistemic stance of the observer, linked to a progression from imitative through analytical to inquiry-oriented observation, and identifies the conditions and open questions of its implementation.
Keywords: *professional vision, classroom observation, practicum, German as a foreign language, teacher education*

Introduction

Pre-service teachers of German in Türkiye spend a substantial part of their final year of study in classrooms. Only a small part of this time is devoted to their own teaching. The remainder consists of a condition that the regulation governing the practicum does not describe, that has no assigned task, and for which the regulatory text has no designation. These hours are the focus of this chapter.

[1] Asst. Prof. Dr., Nevşehir Hacı Bektaş Veli University, Faculty of Education, Department of Foreign Language Education, German Language Teaching Program, ORCID: 0000-0001-8820-8129, *ebruustun@nevsehir.edu.tr*, Türkiye

[2] Assoc. Prof. Dr., Nevşehir Hacı Bektaş Veli University, Faculty of Education, Department of Foreign Language Education, German Language Teaching Program, ORCID: 0000-0002-9490-3109, *bustun@nevsehir.edu.tr*, Türkiye

The Practicum and Its Unregulated Remainder

The practicum is governed by the 2021 regulation of the Ministry of National Education. Article 7/1 sets it at two semesters with six hours of school attendance per week, 144 hours in total, while article 7/2 ties the minimum amount of the pre-service teachers' own teaching to the number of weekly hours allocated to the school subject: at least ten hours per semester if the subject is scheduled for one to two hours per week, and at least twenty hours if it is scheduled for three or more hours per week (Ministry of National Education, 2021).

The arithmetic of these two provisions is the starting point of this chapter. Depending on the subject, the larger part of each semester falls outside the minimum teaching requirement. Because the regulation defines the practicum as the acquisition of teaching competence together with administrative and extracurricular activities (article 4/1-g), not all of this remaining time is spent in the classroom. What the regulation does not do, however, is specify what the pre-service teachers are to do with it. Neither the definition in article 4/1-g nor the task list in article 6/11 assigns an observation task, an observation guide, or a role for this time. Observation appears in the text, but only in one direction: under articles 6/5c and 9/1-ı and in the evaluation form in the annex, it is the university supervisor and the school-based mentor who observe, and the pre-service teachers who are observed (Ministry of National Education, 2021).

This silence is not a regulatory gap. Article 9/1-g requires the university supervisor, school-based mentor, and pre-service teachers to prepare an activity plan together, and article 9/1-ğ requires the pre-service teachers to carry out the activities specified there and report on each of them (Ministry of National Education, 2021). The regulation therefore requires the remaining hours to be planned while leaving the content of that planning entirely open. It delegates rather than omits. Since authority over curricula was transferred to universities in 2020, this delegation has been taken up at the institutional level. The question addressed in this chapter is not whether this time is used, but on what basis it could be structured.

What Unstructured Presence Produces

Recent findings from the Turkish practicum show what happens when this time remains unstructured. Kırmızı et al. (2025) followed 26 pre-service teachers of English through a practicum and conducted weekly joint reflection sessions that were deliberately kept open, without instructions to attend to a particular aspect. The pre-service teachers primarily attended to communication, the physical environment, and classroom management, while the content dimension received comparatively little attention

and was addressed only superficially. The authors also reported that feedback from the school-based mentor often remained superficial and that the intensity of university supervision varied considerably. They therefore call for tools that direct pre-service teachers' attention and for thematically focused rather than open reflection (Kırmızı et al., 2025).

This chapter takes up this call. It asks what such tools would need to look like if they were tailored to German as a foreign language instruction and seeks the reason for their absence in the regulatory design described above. One limitation should be stated at the outset: the study concerns pre-service teachers of English and a training pathway that differs from the 4-year bachelor's degree program in German teacher education. Its findings are treated here as an indication of what unguided presence may produce, not as a description of the German as a foreign language practicum.

Naming The Role and The Aim of the Chapter

Structuring these hours requires naming what happens within them, and the available vocabulary is inconsistent. Three levels must be distinguished. The practicum is an organizational and curricular unit defined by duration, credits, supervision, and assessment; it says nothing about the activity performed within it. Classroom observation is a methodologically regulated activity involving observation categories and a separation between description and interpretation; it is not tied to any institutional setting and can be practiced using video in a university seminar (Gaudin & Chaliès, 2015). Between these two levels lies the role of the person who is present in another teacher's lesson.

German educational terminology names this intermediate level. *Hospitation* derives from the Latin hospes, guest, and defines the role through the status of the person present rather than through that person's activity. Lexical definitions describe the verb as staying in an institution as a guest in order to become familiar with its structures, procedures, and professional problems (Duden.de, 2018). Presence and an intention to learn are thus combined in the word itself, and the term has been extended to video-based settings as virtual *Hospitation* (Stahl, 2023). This chapter defines the role through two implications arising from guest status: no responsibility for conducting the lesson and no authority to evaluate it. The role is unnamed in a sense more consequential than the terminological one. It is unnamed in the regulation governing the practicum, and a condition that the regulation does not name is unlikely to be deliberately structured.

The chapter is a conceptual contribution and did not collect its own empirical data. Its analysis is based on the current regulation, four institutional

practicum documents from Turkish universities, one of which comes from a German teacher education department, and published research on professional vision. The four documents were obtained on the basis of availability rather than through a sampling procedure and make no claim to represent the sector. "*The Institutional Framework of the Practicum in Türkiye*" examines the national and local levels of that framework. "*Observation as a Knowledge-Based Process*" analyzes why observation does not in itself lead to learning and develops noticing as a knowledge-based process. "*Generic and Subject-Specific Professional Vision in the Language Classroom*" takes up the question of subject specificity, and "*A Three-Axis Framework for Structured Observation*" presents the proposed framework. "*Conditions for Implementation*" addresses what its introduction would require, and Limitations and Outlook sets out the limits of the proposal.

The Institutional Framework of the Practicum in Türkiye

Historical Development of the Current Framework

The practicum has been converted into a dual-semester model. Previously, the Higher Education Council divided school practice into *Okul deneyimi* (school experience) and *Öğretmenlik Uygulaması* (teaching practice) (Çam Tosun, 2019; Seyhan Yücel, 2018). It abolished *Okul deneyimi* on the grounds that the course had not been implemented functionally and effectively and integrated its content into *Öğretmenlik Uygulaması*, which was expanded into the sequence *Öğretmenlik Uygulaması 1* and *Öğretmenlik Uygulaması 2* (Yükseköğretim Kurulu, 2018). In 2020, curriculum authority was transferred to the Faculties of Education, which develop and update their programs with reference to the national qualifications framework and the Ministry's teacher competencies (Deregözü, 2020; Çobanoğulları, 2023). A subsequent protocol led to the 2021 regulation, which replaced the 2018 version (Ministry of National Education, 2021).

In the course of this change, student observation ceased to be a credit-bearing component. It survives only as a recommendation: the Higher Education Council notes that the absence of practical hours from the weekly schedule of certain courses does not mean that no practice takes place and that pre-service teachers should be encouraged to conduct observations in schools, classrooms, or laboratories even without receiving credit (Yükseköğretim Kurulu, 2018). This passage concerns courses without formally designated practical hours rather than the practicum course itself. It is also the final mention of observation in the reform volume: a recommendation

without an institutional body responsible for it, without credits, and without relevance to assessment.

Participants, Attendance, and Activity Plan

The 2021 regulation identifies three participants, the pre-service teachers, the school-based mentor, and the university supervisor, and regulates their collaboration through numerical ratios. A school-based mentor may supervise up to six pre-service teachers and must hold the certificate issued by the Ministry. The groups are formed so that no more than twelve pre-service teachers are assigned to one university supervisor (Ministry of National Education, 2021, Madde 4/1-n, 4/1-ğ, 4/1-o, 9/1-a, 9/1-c). The internal regulation of Nevşehir Hacı Bektaş Veli University additionally stipulates that no more than three pre-service teachers may be assigned to a single lesson (Nevşehir Hacı Bektaş Veli University, 2021, Madde 9/1-e). The pre-service teachers teach themselves in at least eight weeks of the semester, while the university supervisor's school visits are tied to these teaching attempts, with at least two on-site observations per semester (Ministry of National Education, 2021, Madde 7/2, 6/5-c). Weekly guidance therefore rests with the school-based mentor, while the university supervisor's presence at the school is limited to assessment visits.

The regulation governs the pre-service teachers' own teaching but leaves the content of the remaining hours to local actors through explicit delegation. Under article 9/1, the university supervisor, school-based mentor, and pre-service teachers prepare an activity plan at the beginning of the semester, and under article 9/1-ğ the pre-service teachers carry out these activities, write a report on each of them, and submit it to the university supervisor (Ministry of National Education, 2021). This obligation is not accompanied by any thematic specification or determination of content.

Local Implementations in Four Documents

How this delegation is taken up can be examined through four documents with differing administrative scopes: the university-wide regulation of Nevşehir Hacı Bektaş Veli University (2021), the faculty guides of Çukurova University (2021) and Bursa Uludağ University (2026), and the departmental guide of the German department at Gazi University (2023), the only document in the corpus prepared for a German department. The corpus is an availability-based selection and makes no claim to representativeness. Administratively, a directive adopted by a university senate differs fundamentally from an implementation-oriented guide. None of the four documents contains a language-related or foreign-language-specific observation category, whether for target language use, bilingual interaction, or language pedagogy, and this also applies to the departmental document.

The Nevşehir regulation provides a student observation form (*Ek 5: Ders Gözlem Formu*) with 11 open-ended guiding questions on planning, lesson opening, objectives, methods, communication, transitions, interaction with learners, lesson closure, and assessment. Its weekly plan (*Ek 3*) assigns lesson planning, implementation, self-evaluation, and portfolio work from week 3 onward, but contains no observation slot and assigns no grade weighting to the form (Nevşehir Hacı Bektaş Veli University 2021). The Çukurova guide limits student observation to week 2 of the activity calendar; its instrument (*Ek-9, Haftalık Değerlendirme Formu*) consists of a header and a single free-text field requiring an evaluation of the week's observation in approximately 200 words, without categories or a guide (Çukurova Üniversitesi, 2021).

The Bursa Uludağ University guide refers to the *Clinical supervision model* for observation and feedback techniques and mentions the five-day Ministry course built around it (Bulunuz& Gürsoy, 2018, as cited in Bursa Uludağ Üniversitesi, 2026). The assignment of school-based mentors is tied to the certificate. The guide provides a calendar of ten possible weeks from which four teaching weeks are selected, together with a self-evaluation sheet after each lesson taught and a peer sheet completed by fellow pre-service teachers, as well as at least two visits by the university supervisor followed by a triadic feedback session. There is no instrument for the systematic observation of the school-based mentor's teaching (Bursa Uludağ Üniversitesi, 2026).

The Gazi guide contains a student observation form (*Form 3: Öğretmen Adayları İçin Öğretmenlik Uygulaması 1 dersi Gözlem Formu*), which is completed once on the basis of the first two weeks and included in the portfolio, which accounts for 20% of the grade. Its 17 items are divided into the physical environment and the educational process and use preschool terminology such as *öğrenme merkezleri*, *gelişim özellikleri*, *aylık plan*, and *kazanım göstergeler*, together with items on individualized support plans. These terms appear alongside the definition of the practicum school as preschool institutions and the requirement for a kindergarten observation sheet in the final portfolio (Gazi Üniversitesi, 2023).

Table 1. Student observation provisions in the four institutional documents

	Nevşehir (2021)	Çukurova (2021)	Bursa Uludağ (2026)	Gazi (2023)
Document type	Regulation	Faculty guide	Faculty guide	Depart-mental guide
Observation instrument	none	Week 2	none	First 2 weeks
Subject-specific categories	none	none	none	none

The Two-Hour Weekly University Seminar

Threee of the documents provide for a two-hour weekly session at the university, while the Uludağ guide does not mention one (Çukurova Üniversitesi, 2021; Gazi Üniversitesi, 2023; Nevşehir Hacı Bektaş Veli Üniversitesi, 2021). The national regulation is silent on this matter. The seminar is therefore documented only locally, not at the national level. The documents relate its content directly to the practicum: in Nevşehir, two theoretical hours accompany the six hours at school, and the university supervisor reviews the development of the school week with the pre-service teachers and discusses the problems they have identified (Nevşehir Hacı Bektaş Veli Üniversitesi, 2021, Madde 12/1, 12/3, 13/1-a, 13/1-d); in Çukurova, the pre-service teachers discuss their weekly reflections and self-evaluation sheets under the guidance of the university supervisor (Çukurova Üniversitesi, 2021); in Gazi, the sessions are organized around lesson planning, microteaching, classroom management, and materials development (Gazi Üniversitesi, 2023). None of the documents limits the seminar to the pre-service teachers' own teaching, and none identifies the observation of others' teaching as its object.

Observation as a Knowledge-Based Process

Placing pre-service teachers in classrooms does not in itself produce professional learning. This section explains why physical presence is insufficient and develops the competence that would be required instead.

Prior Experience as a Filter

Pre-service teachers do not enter the classroom as blank slates. Before their training, they have spent thousands of hours observing teachers from the student perspective, an apprenticeship of observation that produces stable and largely unquestioned beliefs about what teaching is (Lortie,

1975, cited in Çelik & Kahraman, 2021). During the practicum, these beliefs operate as filters through which classroom events are interpreted (Meschede et al., 2017; Mühlum et al., 2023). Where no task structures their attention, pre-service teachers are more likely to notice what corresponds to their own school biography and to overlook or misinterpret what contradicts it, so that unstructured observation confirms the models they bring with them rather than testing those models.

Presence Is Not Perception

The classroom is a fast-moving and multidimensional environment in which many instructional, social, and behavioral events occur simultaneously (Stahnke &Blömeke, 2022). Navigating this stream requires professional vision, the knowledge-based ability to notice and interpret the relevant features of classroom situations. The construct originates in linguistic anthropology, where Goodwin (1994) described the discursive practices through which members of a profession construct and interpret the objects of their professional interest, and it was taken up in teacher education, where it is usually modeled as a two component construct: selective attention, noticing, and knowledge-based reasoning (Gaudin & Chaliès, 2015; Sherin & van Es, 2008; Weger, 2019). Noticing directs attention toward learning-relevant events; knowledge-based reasoning brings professional knowledge to bear in order to describe and explain what has been noticed and assess its consequences (Seidel & Stürmer, 2014; Weger, 2019).

Because noticing is knowledge-dependent, it is not a generic ability to look closely. It is guided by the schemas and scripts in the observer's memory, and the relevant knowledge comprises content knowledge, pedagogical content knowledge, and general pedagogical knowledge (Shulmann, 1986; Steffensky et al., 2015). Steffensky et al. (2015) showed that professional vision of classroom management and professional vision of learning support are empirically separable, although positively related, constructs, and that pre-service teachers with a subject-specific focus in their training performed better on learning support, while no difference emerged for classroom management. Experience operates in the same direction. Experienced teachers observe learners' attention and participation more accurately, approach classroom management preventively, and propose concrete alternatives that support learning considerably more often than novices (Stahnke & Blömeke, 2022).

No Description Without Categories

The transition from looking to systematic description depends on the availability of conceptual categories that make the continuous flow of

classroom interaction segmentable (Manzel et al., 2024; Weger, 2019). If these categories are absent, reflections quickly slide from description into evaluation, and pre-service teachers judge a lesson as successful or unsuccessful according to personal preferences (Mühlum et al., 2023). When pre-service teachers analyze videos without structured prompts, they often interpret the lessons shown as negative examples and scarcely recognize that they represent regular practice with constructive potential for change (Yeşil & Hahn-Laudenberg, 2024). The same surface orientation appears in the Turkish practicum, where content-related aspects are addressed only superficially in open reflection (Kırmızı et al., 2025). The 4 institutional documents examined reinforce this tendency because none provides the subject-specific categories that would direct attention, and one contains no student observation instrument at all.

More practicum time does not solve this problem either. The 2018 reform extended the practicum into a two-semester sequence on the assumption that more hours in the classroom would produce more learning (Yükseköğretim Kurulu, 2018). Research on the learning potential of practica, however, calls for restraint: longer practica do not guarantee continuous competence development, and pre-service teachers without guidance may just as readily adopt less adaptive routines or lose professional motivation (Hascher, 2012). If the duration is extended without a corresponding increase in the quality of guidance, the practicum becomes a setting of missed learning opportunities (Hascher, 2012).

What Structured Observation Requires

Professional vision can nevertheless be developed as early as the initial phase of training if the learning environment is designed accordingly. Video-based settings allow teaching to be analyzed at a distance from the demands of ongoing events (Gaudin & Chaliès, 2015; Seidel & Stürmer, 2014), but merely providing them achieves little. A predefined focus, that is, a lens through which pre-service teachers view a lesson, reduces some of the overload, and instructional supports that direct attention lower the cognitive load of video analysis (Bloomberg et al., 2013). Without theory-based knowledge, novices cannot decode the relevant events at all, and pre-service teachers in early stages of training benefit more from simplified, guided video work than from demanding analytical tools (Gaudin & Chaliès, 2015). Repeated work with identical, structured prompts moves them from description to classification and prediction (Gießler, 2019), whereas repeated use of a video instrument without instructional guidance shows no effects (Seidel & Stürmer, 2014). Human guidance is therefore correspondingly important: discussions led by the university supervisor or school-based mentors help pre-service teachers link what they have

seen with theoretical principles (Tripp &Rich, 2012, cited in Gaudin & Chaliès, 2015).

The Clinical supervision model *in the Turkish context*

Türkiye already has a structured model for mediating school-based practicum experiences. The *Clinical supervision model* was introduced to address the low quality and frequency of practicum feedback and organizes the practicum as a collaborative partnership among pre-service teachers, the school-based mentor, and the university supervisor (Gürsoy et al., 2016). The university supervisor coordinates the conferences and observes non-evaluatively, while the school-based mentor changes from an evaluator into a "model, mentor, colleague and coach" (Kiraz & Yıldırım, 2007, p. 250, cited in Gürsoy et al., 2013). The cycle comprises 5 phases: pre-conference, observation and data collection, data analysis, post-conference, and reflection, with the observation focus selected by the pre-service teachers during the pre-conference and the triadic post-conference serving joint analysis and action planning (Bulunuz et al., 2014). School-based mentors are prepared in a 5-day Ministry course based on a dedicated practice guide (Bulunuz & Gürsoy, 2018, cited in Bursa Uludağ Üniversitesi, 2026).

The evaluations report favorable results. Gürsoy et al.(2013) found significantly higher ratings of supervisory performance in the experimental group, including for regular feedback and a positive atmosphere; Bulunuz et al. (2014) found higher performance scores in the analysis of videotaped lessons; Gürsoy et al. (2016) found significantly higher satisfaction with supervision among both pre-service teachers and school-based mentors, while 8 of 18 pre-service teachers in the control group reported that they had never been observed. Throughout, however, the cycle is described exclusively with regard to lessons taught by the pre-service teachers themselves. The model does not structure the systematic observation of the school-based mentors' regular teaching and thus leaves the remaining attendance hours without the pre-service teachers' own teaching unaddressed.

Generic and subject-specific professional vision in language teaching

Teaching quality has long been studied using generic frameworks that capture cross-subject dimensions such as classroom management, cognitive activation, and a supportive climate (Praetorius et al., 2020; Steffensky et al., 2015). Cross-subject research across subject didactics that uses a synthesis framework with 7 quality dimensions as a basis for comparison suggests that

although the highest level of abstraction may be regarded as generic, the subdimensions and, in particular, the observable indicators require considerable subject-specific adaptations; even classroom management exhibits a meaningful degree of subject specificity at the indicator level (Praetorius et al., 2020). The 2-dimensionality of professional vision reported in *Observation as a Knowledge-Based Process* points in the same direction: it is not a unitary ability but a competence that depends on the quality focus under consideration and the knowledge base activated in the process (Steffensky et al., 2015).

Findings from other subjects

The domain-specific character of professional vision has been investigated in several subjects. In science, master's students with a subject-specific focus outperformed the others in the perception of learning support, while no difference emerged for classroom management (Steffensky et al., 2015), and a video-based university teaching intervention improved both professional vision and self-efficacy beliefs concerning language-sensitive teaching in primary science education (Grewe et al., 2023). In mathematics, Star and Strickland (2008) found that pre-service teachers did not enter their methods course with well-developed observation skills, with the weakest initial noticing occurring for features of the classroom environment and mathematical content; after a seminar organized around a dedicated observation framework, noticing improved significantly in these areas and in teacher-student communication. In civic education, the analysis of central practices such as the quality of lesson openings, the facilitation of controversial discussions, and the promotion of political judgment requires specialized categories (Manzel et al., 2024), and without structured categories pre-service teachers slide from description into moral evaluation and label regular practice as a negative example (Yeşil &Han- Laudenberg, 2024).

Initial findings on foreign language teaching

Research on foreign language teaching is more recent and concentrated on English. Uličná (2017) compared the unstructured reflections of Czech pre-service teachers of English with an expert analysis of the same recorded lesson. The experts identified 6 subject-specific moments: a grammar-related warm-up activity, the teacher's code-switching, the teacher's linguistic accuracy, the teacher's speaking time, the inductive teaching of a grammatical phenomenon, and transfer to real language use. The 47 participants commented on an average of approximately 2 of these, and their reflections were predominantly descriptive or evaluative and rarely linked what they had observed to theory. Janík et al. (2021) examined professional vision of student engagement in video-based interviews and found that pre-service teachers observed engagement at the behavioral,

cognitive, and emotional levels but verbalized it in everyday language rather than specialist terminology. Gießler (2019) examined professional vision of lexical learning and showed that open prompts leave room for subjective interpretations, whereas structured, category-guided prompts direct attention toward specific features of language pedagogy such as error correction and explanations of meaning.

The open question for German as a foreign language

For German, the question remains largely open. Professional vision is regarded as a key construct in German teacher education, but systematic, category-guided empirical studies are rare (Schramm & Bechtel, 2019). An exploratory study underscores this point: using the video-based DAF-VideoWeb platform, Janík And Janíková (2019) analyzed the professional vision of 5 pre-service teachers of German and found that their comments were dominated by description, while explanation and prediction remained rare and attention was directed much more strongly toward the teacher than toward the learners. Weger (2019) defines professional vision as the link between a teacher's competence and action and emphasizes the distinctive nature of the foreign language lesson, in which the target language is simultaneously the medium and the object of instruction. Research on language-sensitive teaching points in the same direction: Stahl (2023) develops a university teaching format combining classroom video and observation supports and argues that the perception of language-sensitive quality does not develop on its own but requires explicit support. Mühlum et al. (2023) document a staged sequence of tasks in which descriptions remain as value-neutral as possible and evaluations become increasingly conceptually grounded, while also noting that where tasks leave observation unstructured, personal beliefs about good teaching quickly enter the analysis. This leads to a framework that structures observation in a subject-specific way.

A three-axis model for structured classroom observation

The framework proposed here does not replace existing models of active teaching practice. It is a scaffold for the remaining practicum hours without the pre-service teachers' own teaching and interweaves 3 axes: the object of perception, the form of mediation, and the epistemic stance of the observer.

Axis 1: Object of perception

Axis 1 determines what pre-service teachers attend to. Without explicit subject-specific categories, novices are easily overwhelmed by the

multidimensionality of classroom interaction, and generic quality dimensions do not provide what language teaching requires. Axis 1 therefore groups the object of perception into 5 subject-specific categories: target language use, lexical work, instructional language, interactional and communicative support, and language-learning feedback. Combining these 5 into a coherent framework is a proposal by the authors, whereas the individual aspects on which it draws are established in language teaching research: target language use and teacher speaking time as subject-specific moments of the lesson (Uličná, 2017), professional vision of lexical learning (GIEßLER, 2019), observation of language-sensitive teaching and micro-scaffolding (Stahl, 2023), and the dual status of the target language as both the medium and the object of instruction (WEGER, 2019). What the categories encompass operationally and how they could be translated into observation items lies beyond the scope of this chapter. The central point here is the principle that the object of perception in language teaching cannot be derived from generic quality dimensions alone.

Axis 2: Form of mediation

Axis 2 determines how observation is supported, from unstructured presence to mediated, collaborative reflection. The continuum begins with the traditional guest role, in which pre-service teachers spend dozens of hours in the classroom without tasks or prompts. The framework replaces this condition with mediated observation, in which written supports direct selective attention toward a limited number of events and thereby reduce cognitive load (Blomberg et al., 2013; Gießler, 2019); designing such supports is a task in its own right and is not developed here. Beyond individual prompts, the highest level of mediation lies in dialogic post-observation discussion, in which the observation is taken up in conversation with fellow pre-service teachers, school-based mentors, and the university supervisor (Mühlum et al., 2023).

This axis is designed to connect with the triadic conference structure of the *Clinical supervision model* described in *Observation as a Knowledge-Based Process*. Because its cycle is described exclusively with regard to lessons taught by the pre-service teachers themselves, the hours without their own teaching remain unaddressed, and this is precisely where the present framework extends the model. Structured observation prompts as the basis for discussions in the 2-hour weekly university seminar (Çukurova Üniversitesi, 2021; Gazi Üniversitesi, 2023; Nevşehir Hacı Bektaş Veli Üniversitesi, 2021) would create a collaborative space in which the university supervisor and school-based mentors support pre-service teachers in linking observations with theory. The purpose is to transform classroom presence into a deliberate, goal-directed activity rather than an unguided condition.

Axis 3: Epistemic stance

Axis 3 determines the analytical stance of the pre-service teachers. Without categories and analytical tasks, they slide almost immediately into evaluation and judge a lesson according to their own preferences rather than asking why an interaction took the form observed (Mühlum et al., 2023; Yeşil & Hahn-Laudenberg, 2024). To counteract this drift, we distinguish 3 stances. The distinction is our own; it was developed with reference to the staged typologies of reflection established in teacher education research, which likewise move from largely descriptive accounts toward reflection that weighs alternative explanations and questions what is taken for granted (Hatton & Smith, 1995). Premature evaluation is not regarded here as the first stage of the progression but as an untrained starting state that the progression is intended to displace.

The first stance is descriptive: the observer reconstructs what occurred, uses specialist terminology and the categories of axis 1, and suspends judgment about quality. The second is explanatory: the observer asks why an interaction took the form observed, and what has been noticed is linked to pedagogical concepts and the probable consequences for the learners. The third is inquiry-oriented: the classroom is regarded neither as a repertoire of finished practices to be imitated nor as a performance to be graded, but as a place of open questions in which the observer formulates their own questions, analyzes situations as cases of complex educational problems, and develops reasoned alternatives.

Progression and institutional embedding

The 3 axes are translated into a progression of 3 stages, an imitative, an analytical, and an inquiry-oriented stage, situated within the 2 consecutive practicum courses *Öğretmenlik Uygulaması 1* and *Öğretmenlik Uygulaması 2*.

The imitative stage encompasses the first 2 weeks and corresponds to the introductory observation phase provided for in the departmental guides: the Gazi observation sheet is to be completed on the basis of the observations conducted by the pre-service teachers during their first 2 weeks (Gazi Üniversitesi, 2023, *Form 3*). All 3 axes are at their narrowest values here: a narrow, readily visible object of perception, a strongly guided form of mediation, and a strictly descriptive stance. The focus is on orientation, not yet on analysis. The analytical stage encompasses the following weeks. It fits into the institutional weekly plans, which from week 3 onward require pre-service teachers to prepare a daily plan together with the necessary materials, carry out the planned activities, and prepare a self-evaluation report for their portfolio (Nevşehir Hacı Bektaş Veli Üniversitesi, 2021, *Ek 3*). The object of perception shifts toward the subject-specific categories, mediation becomes semi-structured, and the stance

moves from description to explanation, with the weekly seminar serving as the setting for joint analysis. The inquiry-oriented stage occurs in the second practicum course. The pre-service teachers select their own focus, mediation shifts from external guidance toward collaborative and largely self-directed forms, and the stance becomes inquiry-oriented. The direction of progression is the same across all 3 axes: from narrow to self-selected objects, from strong to lighter mediation, and from description to question.

Through this structuring, the framework aims to give the guest role a defined cognitive task rather than adding further hours to an already dense practicum. Whether it succeeds remains an empirical question because the framework has not yet been implemented or tested.

Conditions for implementation

Institutional conditions

The delegation described in "*The Institutional Framework of the Practicum in Türkiye*" is the starting point for the framework because the regulation requires the remaining hours to be planned without prescribing their content (Ministry of National Education, 2021, Madde 9/1-g, 9/1-ğ). The local situation is inconsistent. 3 of the 4 guides provide for a 2-hour weekly theoretical session in which developments and problems from the practicum school are discussed under the direction of the university supervisor, and none limits this session to the pre-service teachers' own teaching (Çukurova Üniversitesi, 2021; Gazi Üniversitesi, 2023; Nevşehir Hacı Bektaş Veli Üniversitesi, 2021); the Uludağ guide does not mention such a session, although this silence concerns the document and not necessarily the practice it describes. What is missing is not the format but its object.

The situation differs with regard to the instruments. The observation forms and weekly plans are not designed to support subject-specific noticing, and the weeks from week 3 onward are assigned to planning, the pre-service teachers' own teaching, and self-evaluation, while observation of others' teaching remains unnamed (Nevşehir Hacı Bektaş Veli Üniversitesi, 2021). Because these weeks are already institutionally structured, they could accommodate observation tasks without additional practicum hours. The transfer of curricular authority to the faculties in 2020 operates in both directions (Deregözü, 2020): it gives the faculties the latitude to integrate such a framework, while the absence of shared, language-specific specifications makes coordinated implementation more difficult.

Requirements for the participants

The framework requires school-based mentors and university supervisors to shift from evaluative to developmental guidance, which is particularly demanding for school-based mentors who are accustomed to acting as external evaluators rather than as partners who make their own thinking about teaching visible. The Ministry's certificate course is the established route for this qualification (Ministry of National Education, 2021, Madde 4/1-ğ), and the Uludağ guide identifies it as a 5-day course based on the *Clinical supervision model* (Bursa Uludağ Üniversitesi, 2026). Because this model is organized around lessons taught by the pre-service teachers themselves, school-based mentors are not currently prepared to guide the systematic observation of regular teaching; an extension of the existing course would be required.

The university supervisor faces separate constraints. The groups are formed so that no more than 12 pre-service teachers are assigned to one university supervisor (Ministry of National Education, 2021, Madde 4/1-o, 9/1-a,9/1-c), often distributed across several schools, which makes regular visits and sustained dialogic feedback difficult.

Media and technological options

Video-based settings can mitigate some of these constraints by decoupling observation from physical presence, as demonstrated by the Viko_spracHe corpus and the DAF-VideoWeb platform (Janík & Janíková, 2019; Stahl, 2023). Recordings can be paused and viewed repeatedly, reducing the immediate cognitive demand. The limitations are equally clear. Without structured prompts and observation supports, pre-service teachers are overwhelmed by the density of what they see or attend only to issues of classroom management (Blomberg et al., 2013; Gaudin & Chaliès, 2015). Producing and distributing classroom recordings requires consent and raises data protection issues, and platforms presuppose infrastructure and digital competence that are not equally available everywhere. Video preparation and digital tasks can support the early analytical stages (Mühlum et al., 2023), but they do not replace the situated experience of the classroom and should be understood as a supplement.

Open questions

The framework is a conceptual proposal, and its value depends on questions that only empirical work can answer. The first concerns the object of perception: whether subject-specific categories of this kind are accessible to pre-service teachers at this stage of training and whether they change what the pre-service teachers attend to. The second concerns the

epistemic stance: whether a progression from description through explanation to question can withstand the documented drift into premature evaluation, and under what conditions. The third concerns feasibility: what such an approach would require of school-based mentors and the university supervisor under the constraints described. The fourth concerns durability: whether anything acquired during the practicum carries over into the pre-service teachers' later teaching, the question on which the educational claim of structured observation ultimately depends.

Limitations and Outlook

Limitations

This proposal has 4 limitations. The first is that the framework is conceptual and untested, so nothing can yet be said about its effect on professional vision, its manageability for university supervisors, or its feasibility for school-based mentors under real conditions. The second is the narrow document base. The corpus was an availability-based selection of 4 documents with differing administrative scopes: 1 university-wide regulation, 2 faculty guides, and 1 departmental guide, the latter being the only document prepared for a German department. This selection cannot claim representativeness for the diverse field of German teacher education in Türkiye. The third is that the institutional analysis is based on written documents rather than observed practice; guides describe formal structures but not how the practicum is negotiated in everyday practice, where informal routines and adaptations may differ considerably from the written requirements. The fourth is the narrow empirical base in German as a foreign language on which the subject-specific categories are based; several of the relevant studies use very small samples, including the exploratory study by JANÍK and JANÍKOVÁ (2019) with 5 participants.

Outlook

The framework does not require restructuring the national practicum system. Its principal setting is the 2-hour weekly seminar that 3 of the guides examined already provide for and that none limits to the pre-service teachers' own teaching (Çukurova Üniversitesi, 2021; Gazi Üniversitesi, 2023; Nevşehir Hacı Bektaş Veli Üniversitesi, 2021). What would be added there is not a format but an object, namely the categories through which others' teaching could be observed and discussed. The Ministry's certificate course provides a comparable channel for the qualification of school-based mentors (Ministry of National Education, 2021, Madde 4/1-ğ) which, if extended to include guidance of pre-service teachers' perception during *Hospitation* hours, would allow them to act as facilitators of

observation rather than as evaluators of teaching. For research, the framework provides a starting point for examining how the transition from premature evaluation to descriptive, explanatory, and inquiry-oriented stances unfolds under different forms of mediation.

Taken together, these proposals aim to shift the emphasis of the practicum away from early teaching attempts alone and toward a connection between observation and action. The development of professional vision is not a passive by-product of practicum attendance but a prerequisite for pedagogical action. Teacher education must first ensure that pre-service teachers learn to see before they are expected to learn to teach.

Declaration on the use of artificial intelligence

AI-based tools were used to organize the source corpus, for language editing and for support in translation. All resulting content was checked against the primary sources by the authors. The framework presented here and aşş interpretations are the authors' own and the authors take full responsibility for the content of this chapter.

References

Blomberg, G., Renkl, A., Sherin, M. G., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online, 5*(1), 90–114.

Bulunuz, N., Gürsoy, E., Kesner, J., Baltacı Göktalay, Ş., & Salihoğlu, U. M. (2014). The Implementation and Evaluation of a Clinical Supervision Model in Teacher Education in Türkiye: Is It an Effective Method? *Educational Sciences: Theory & Practice, 14*(5), 1823–1833. https://doi.org/10.12738/estp.2014.5.2085

Bursa Uludağ Üniversitesi. (2026). *Öğretmenlik Uygulamasi Uygulama Esaslari | Eğitim Fakültesi | Bursa Uludağ Üniversitesi*. Uludag.Edu.Tr. https://uludag.edu.tr/egitim/ogretmenlik-uygulamasi-uygulama-esaslari-27510

Çam Tosun, F. (2019). Öğretmenlik Uygulamasının Süre Açısından İncelenmesi ve Bir Çözüm Önerisi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 52*(3), 839–869. https://doi.org/10.30964/auebfd.539374

Çelik, O. T., & Kahraman, Ü. (2021). Göreve Yeni Başlayan öğretmenlerin Yaşadıkları Güçlükler. *Pamukkale University Journal of Education*, (51). https://doi.org/10.9779/pauefd.684913

Çobanoğulları, F. (2023). Almanya'daki Yabancı Dil Öğretmen Adaylarının Öğretmenlik Uygulamaları Eğitimi. *Sosyal Bilimler Enstitüsü Dergisi, 13*(2), 971–981. https://doi.org/10.30783/nevsosbilen.1257757

Çukurova Üniversitesi. (2021). *Öğretmenlik Uygulaması Kılavuzu*. Eğitim Fakültesi. https://egitim.cu.edu.tr/storage/uygulama/UygulamaKilavuzu.pdf

Deregözü, A. (2020). Geçmişten günümüze almanca öğretmenliği lisans programlarına İlişkin karşılaştırmalı bir çalışma. *Diyalog*, 143–161. https://doi.org/10.37583/diyalog.802258

Duden.de. (2018, April 27). *Hospitation*. Duden.de. Duden. https://www.duden.de/rechtschreibung/Hospitation

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41–67. https://doi.org/10.1016/j.edurev.2015.06.001

Gazi Üniversitesi. (2023). *2022-2023 Eğitim Öğretim Yılı Öğretmenlik Uygulaması 1 Ve 2 Dersi Kılavuzu* . https://webupload.gazi.edu.tr/upload/88/2024/6/4/9826dbc9-a99c-4ff7-b132-110f63bf0386-alman-dili-ve-eg-itimi-abd-kilavuz-1.pdf

Gießler, R. (2019). Schriftliche Aufgabenformate zur Erfassung und Förderung der professionellen Unterrichtswahrnehmung von angehenden Englischlehrpersonen. *Fremdsprachen Lehren Und Lernen, 48*(1), 32–49. https://doi.org/10.2357/FLuL-2019-0003

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Grewe, O., Zucker, V., Todorova, M., Möller, K., & Meschede, N. (2023). Förderung Der Professionellen Unterrichtswahrnehmung und der selbstwirksamkeitsüberzeugungen von masterstudierenden bzgl. Sprachsensibler maßnahmen im Naturwissenschaftlichen Sachunterricht. *Zeitschrift Für Didaktik Der Naturwissenschaften, 29*(1). https://doi.org/10.1007/s40573-023-00163-y

Gürsoy, E., Bulunuz, N., Baltacı Goktalay, S., Bulunuz, M., Kesner, J., & Salihoğlu, U. (2013). Clinical Supervision Model to Improve Supervisory Skills of Cooperating Teachers and University Supervisors during Teaching Practice. *H. U. Journal of Education, (1*), 191–203.

Gürsoy, E., Kesner, J. E., & Salihoglu, U. M. (2016). Clinical Supervision Model in Teaching Practice: Does it Make a Difference in Supervisors' Performance? *Australian Journal of Teacher Education, 41*(11), 61–76. https://doi.org/10.14221/ajte.2016v41n11.5

Hascher, T. (2012). Lernfeld Praktikum – Evidenzbasierte Entwicklungen in der Lehrer/innenbildung. *Zeitschrift Für Bildungsforschung, (2)*, 109–129. https://doi.org/10.1007/s35834-012-0032-6

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education, 11*(1), 33–49. https://doi.org/10.1016/0742-051x(94)00012-u

Janík, M., & Janíková, V. (2019). Entwicklung der professionellen Wahrnehmung künftiger DaF-Lehrer(innen) mittels einer videobasierten Online-Plattform (DaF-VideoWeb). *Fremdsprachen Lehren Und Lernen, 48*(1), 63–83. https://doi.org/10.2357/FLuL-2019-0005

Janík, M., Minaříková, E., Janík, T., & Juříková, Z. (2021). On the Language of Student Teachers' Professional Vision: How Do Pre-Service EFL Teachers Comment on Classroom Videos of Pupil Engagement? *Orbis Scholae, 15*(3), 43−60. https://doi.org/10.14712/23363177.2022.14

Kırmızı, Ö., Irgatoğlu, A., & Bayındır Özkan, F. Y. (2025). Pre-service EFL teachers' noticing in practicum through joint reflection in the Turkish context. *Sage Open, 15*(2), 1–13. https://doi.org/10.1177/21582440251342337

Manzel, S., Gronostay, D., & Hahn-Laudenberg, K. (2024). Professionelle Unterrichtswahrnehmung von Politiklehrkräften. In D. Gronostay, S. Manzel, K. Hahn-Laudenberg, & J. Teuwsen (Eds.), *Professionelle Unterrichtswahrnehmung im sozialwissenschaftlichen Fachunterricht Das Potential von Animationsfilmen realer Unterrichtsszenen für die Lehrer*innenbildung* (pp. 3–19). SpringerVS.

Meschede, N., Fiebranz, A., Möller, K., & Steffensky, M. (2017). Teachers" professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education, 66*, 158–170. https://doi.org/10.1016/j.tate.2017.04.010

Ministry of National Education. (2021). *Öğretmenlik Uygulaması Yönergesi.* Millî Eğitim Bakanlığı Tebliğler Dergisi. https://akademi.meb.gov.tr/www/ogretmenlik-uygulamasi-yonergesi/icerik/71

Mühlum, E., Saunders, C., & Sternberg, J. (2023). Unterrichtsgeschehen wahrnehmen Die Schulung professioneller Unterrichtswahrnehmung mittels Videomitschnitten (UDOs) in DLL. *KONTEXTE - Internationales Journal Zur Professionalisierung in Deutsch Als Fremdsprache, 1*(2), 46–63. https://doi.org/0.24403/jp.1335304

Nevşehir Hacı Bektaş Veli University. (2021). *Lisans Programları Öğretmenlik Uygulaması Dersi Yönergesi*. https://sistem.nevsehir.edu.tr/yonerge-politika/38c043d38b8096cae018493fe9b1aa53.pdf

Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., & Nehring, A. (2020). Unterrichtsqualität in den fachdidaktiken im deutschsprachigen raum – zwischen generik und fachspezifik. *Unterrichtswissenschaft, 48*(3), 409–446. https://doi.org/10.1007/s42010-020-00082-8

Schramm, K., & Bechtel, M. (2019). Videobasierte Lehre in der Fremdsprachendidaktik. *Fremdsprachen Lehren Und Lernen, 48*(1), 3–13. https://doi.org/10.2357/FLuL-2019-0001

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal, 51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Seyhan Yücel, M. (2018). Ansichten von Lehramtskandidaten und Lehrkräften der Deutschlehrerausbildung im Hinblick auf das Schulpraktikum am Beispiel der Universität Trakya. *Diyalog,* (2), 169–182.

Sherin, M. G., & van Es, E. A. (2008). Effects of video club participation on teachers" professional vision. *Journal of Teacher Education, 60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.2307/1175860

Stahl, C. (2023). Sprachsensiblen Unterricht professionell wahrnehmen – ein Lehrkonzept mit Aufgabenformaten zur Entwicklung professioneller Unterrichtswahrnehmung. *KONTEXTE - Internationales Journal Zur Professionalisierung in Deutsch Als Fremdsprache, 1*(1), 101–119. https://doi.org/10.24403/jp.1296789

Stahnke, R., & Blömeke, S. (2022). Wie erfahrene und angehende lehrkräfte unterricht wahrnehmen. *Journal Für lehrerInnenbildung Jlb 22. Jahrgang (2022) Heft 4 Classroom Management*, (4), 76–85. https://doi.org/10.35468/jlb-04-2022-04

Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*(2), 107–125. https://doi.org/10.1007/s10857-007-9063-7

Steffensky, M., Gold, B., Holdynski, M., & Möller, K. (2015). Professional vision of classroom management and learning support in science classrooms—Does professional vision differ across general and content-specific classroom interactions? *International Journal of Science and Mathematics Education, 14*(5), 1001–1002. https://doi.org/10.1007/s10763-015-9627-4

Uličná, K. (2017). Professional vision of future English language teachers: subject-specific noticing and knowledge-based reasoning. *E-Pedagogium*, (11), 38–50.

Weger, D. (2019). Professional Vision – State of the art zum Konstrukt der professionellen Unterrichtswahrnehmung in der Lehrer- (innen)bildung. *FLuL - Fremdsprachen Lehren Und Lernen, 48*(1), 14–31. https://doi.org/10.2357/FLuL-2019-0002

Yeşil, K., & Hahn-Laudenberg, K. (2024). Kontroversität und Diskursivität im Unterricht. Integration von LArS-Modulen in bildungswissenschaftliche Lehrveranstaltungen unter demokratiepädagogischen Fragestellungen . In D. Gronostay, S. Manzel, K. Hahn-Laudenberg, & J. Teuwsen (Eds.), *Professionelle Unterrichtswahrnehmung im sozialwissenschaftlichen Fachunterricht Das Potential von Animationsfilmen realer Unterrichtsszenen für die Lehrer*innenbildung* (pp. 199–226). SpringerVS.

Yükseköğretim Kurulu. (2018). *Öğretmen Yetiştirme Lisans Programları*. Yükseköğretim Kurulu. https://egitim.yok.gov.tr/tr/document/2455

Motivation to Learn with GenAI: A Self-Determination Theory Perspective on Pre-Service German Language Teachers[1]

Hamdullah Şahin [2]

Abstract
The widespread adoption of generative artificial intelligence (GenAI) offers distinct advantages for language education. But its sustainable integration depends heavily on student motivation. Although existing literature mostly explores this dynamic in general and in English learning contexts, the motivational profiles of pre-service German language teachers remain largely unexamined. Based on self-determination theory (SDT), this correlational survey study investigates the motivation of 94 pre-service German language teachers regarding GenAI use. The data was collected using the AI motivation scale (AIMS) and analyzed through descriptive statistics and correlation analysis. The findings demonstrated that participants—who primarily utilized tools like ChatGPT and Gemini—exhibited high levels of autonomous motivation (intrinsic and identified regulation) alongside notably low levels of controlled motivation (external and introjected regulation) and amotivation. A strong negative correlation between intrinsic motivation and amotivation indicates that genuine interest, rather than external pressure or regulatory factors, drives GenAI adoption among teacher candidates. Consequently, the results imply that integrating GenAI into teacher training programs necessitates a hybrid model that accounts for both intrinsic and extrinsic motivational needs, relying on teacher guidance to support autonomous learning.
Keywords: generative artificial intelligence, German language teaching, motivation, pre-service teachers, self-determination theory

Introduction

The development and widespread adoption of generative artificial intelligence (GenAI) technologies have fundamentally transformed language education. Through individualized, real-time feedback, personalized content generation, and flexible practice opportunities, these systems offer considerable potential for both learners and teachers to improve the

[1] An earlier and shorter version of this study was presented as a conference paper at the 11th International Conference on Education and Teaching (ICETL), Helsinki, July 2026. This chapter presents an expanded and revised version with substantial changes.

[2] Asst. Prof. Dr., Anadolu University, Faculty of Education, Department of Foreign Language Education, German Language Teaching Program, ORCID: 0000-0002-1575-1611, *hamdullahsahin@anadolu.edu.tr*, Türkiye

quality of language education (Cordero et al., 2025). With these systems, learners can set their own pace, identify their own weaknesses, receive feedback on them more quickly than in traditional instruction, and achieve a more autonomous learning environment (Creely, 2024).

However, the adoption of GenAI in education cannot be viewed exclusively as a positive development. The concerns regarding GenAI-facilitated plagiarism have become a primary topic of debate in recent years (Nelson et al., 2025). Similarly, researchers have identified various risks, especially about students, such as developing dependency on GenAI and consequently losing essential cognitive skills (Creely, 2024). Other risks highlighted are that the systems produce hallucinations which decrease educational quality (Nadim & Di Fuccio, 2025), the potential transmission of biases and a possible digital divide resulting from unequal access to GenAI (Cordero et al., 2025), and broader concerns regarding social isolation and data security (Aly et al., 2025).

To address such issues and provide a high-quality educational environment, motivation has emerged as a critical factor in navigating the widespread adoption of GenAI in language education (Ahmed & Sultan, 2026; Demirci et al., 2025). Particularly in learning environments such as language education, where socialization and human interaction are prioritized, the inappropriate use of GenAI can force students into a passive role (Creely, 2024). This prevents them from going through the stages of sustainable learning, thereby hindering the long-term retention of the learned material (Creely, 2024). In such cases, while students may feel they learn autonomously, they might actually be blinded by a false sense of achievement created by GenAI (Wang et al., 2025). When students attempt to use the learned language in an authentic setting or are required to engage with more complex academic tasks, they may suddenly confront their own shortcomings, lose their self-confidence, and experience a decline in motivation (Riaz et al., 2026).

Therefore, to establish a sustainable learning environment and minimize the potential risks of GenAI, the structure of motivation and its relationship with GenAI in language education need to be clarified. In this regard, Ryan and Deci's (2000) self-determination theory (SDT) offers a robust theoretical framework for explaining motivation. Focusing on individuals' basic psychological needs for autonomy, competence, and relatedness, SDT examines motivation across a spectrum of intrinsic motivation, extrinsic motivation, and amotivation (Ryan & Deci, 2020). SDT explains the conditions under which each type of motivation tends to develop or conversely, diminish. Over the years, SDT has been widely applied to understand students' learning motivation, and more recently, it has also been

used to examine how the integration of technology into education affects motivation (Demirci et al., 2025).

Considering this, the primary aim of the current study is to examine the motivation of pre-service German language teachers regarding learning with GenAI technologies within the framework of SDT. Firstly, the concept of motivation was explained with a central focus on SDT, and its connection to GenAI-supported language education was presented in light of relevant studies in the present study. Then, the empirical section reported the results of the AI motivation scale administered to 94 pre-service German language teachers to assess their motivation to learn with GenAI. Descriptive statistics and correlation analysis were used to analyze the scale responses, and the results were discussed in relation to both SDT and GenAI-supported language education. Finally, drawing on the theoretical background and the empirical findings, this study offered recommendations for the effective use of GenAI and how to strengthen student motivation regarding its use in German language teaching and learning.

Definitions of Motivation and Self-Determination Theory

Motivation is generally defined as a complex construct encompassing cognitive, emotional, and social dimensions (Lukianenko, 2025). It is the driving force behind people's willingness to act and devote time to a given subject or activity, and also sustains individuals' engagement with that activity over the course of their lives (Espinar Redonda & Ortega Martin, 2015). It is also described as a multilayered construct containing intrinsic and extrinsic motivators, self-efficacy, self-regulation, and the learning environment (Moybeka et al., 2023).

For consistent and sustainable learning to occur, students must first have the desire to learn. Therefore, motivation is acknowledged to be an integral part of a successful and effective learning process (Fan et al., 2024). To illustrate, motivation directly influences students' academic achievement and persistence in learning, their desire to reach learning goals, and their resilience in overcoming obstacles encountered during this process (Chiu et al., 2024). It is also widely known that highly motivated students engage more actively in the learning process, both behaviorally and cognitively and thus, they internalize this process to a greater extent, which makes learning more enduring (Li et al., 2026). In line with this, motivation promotes students' independent learning and self-regulation skills. Consequently, through the development of autonomous learning skills, students can acquire lifelong learning habits (Fan et al., 2024).

When all these aspects of motivation are considered in the context of language education, it appears that motivation is the underlying force driving students to learn a new language and to acquire proficiency in that language (Lukianenko, 2025). Students need motivation to achieve their linguistic goals, maintain their overall interest in the language, dedicate time to necessary assignments and practice, and cognitively prepare themselves for this process (Espinar Redonda & Ortega Martin, 2015). Fundamentally, language education involves learning about a different culture, a new country, and interacting with its people. Using the language correctly, achieving these learning goals, and internalizing the process, in turn, positively affect motivation (Wang et al., 2025).

Concerning the theoretical perspectives of motivation, Gardner and Lambert (1972) distinguish between *integrative* and *instrumental motivation* in their framework of second language acquisition. They define integrative motivation as a language learner's emotional affinity for the target language, their desire to become part of the culture where it is spoken, their appreciation for its people and values, or their desire to live, work, or form relationships within that community. Instrumental motivation, on the other hand, describes learners acquiring the language not for their personal interest, but for academic success, self-improvement, or general necessity.

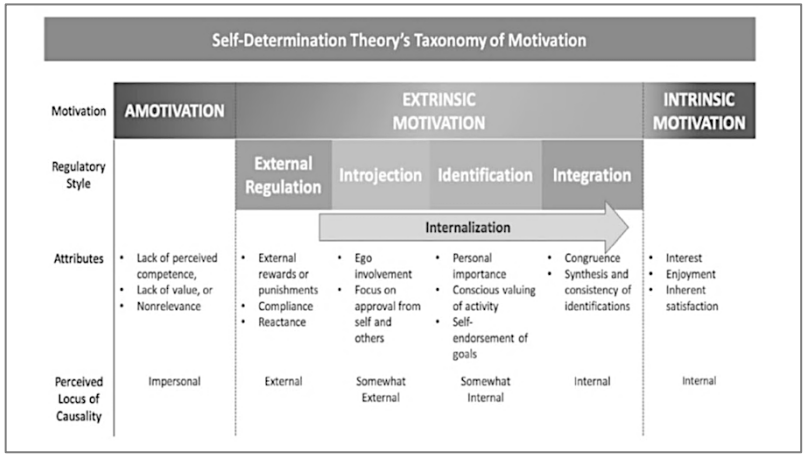
Gardner and Lambert's (1972) framework was further expanded by Dörnyei (1994). Dörnyei (1994) groups Gardner and Lambert's integrative and instrumental motivation under another sub-dimension called the *language level*. Accordingly, there are two other important sub-dimensions for motivation in foreign language teaching: the *learner level* and the *learning situation level*. Among these sub-dimensions, the learner level encompasses personal characteristics and cognitive processes such as the individual's need for achievement, self-confidence, anxiety level, and beliefs about their own abilities. The learning situation level, however, is directly related to factors such as the classroom environment, curriculum and materials, teachers' approaches and feedback as well as the overall cohesion, commitment, and goals of the class.

A more comprehensive and generalized form of these motivation theories in foreign language learning can be found in Ryan and Deci's (2000) SDT. Further developed by Ryan and Deci over the years and now widely accepted in the academic world, this theory holds that every individual has three basic psychological needs that must be met for natural development, success, and intrinsic motivation to occur: *autonomy*, *competence*, and *relatedness* (Ryan & Deci, 2000). *Autonomy* refers to individuals' ability to initiate their own actions, make their own choices, and maintain control

over their behavior (Ryan & Deci, 2000). In the context of education, this includes students making their own decisions and ensuring their behavior aligns with their own interests and values (Chiu et al., 2024). *Competence* is an individual's sense of succeeding at a task and making effective progress (Ryan & Deci, 2000). It can also be defined as possessing the knowledge and skills needed to carry out a complex task successfully (Shi & Zhang, 2025). The final need, *relatedness*, concerns being connected with other students and teachers during the learning process and feeling like part of a social structure (Ryan & Deci, 2000).

SDT posits that in situations where these three needs are supported, motivation would become more internalized and enduring, thereby making it more likely that educational goals would be achieved. Nevertheless, according to Ryan and Deci (2000), not all types of motivation are equal. From an educational perspective, the type of motivation a person has can be more important than how much they are motivated. While Ryan and Deci (2020) categorize motivation into three main types — intrinsic motivation, extrinsic motivation, and amotivation, they also explain their sub-dimensions and meanings within the SDT taxonomy (see Figure 1).

Figure 1: *SDT Taxonomy by Ryan and Deci (2020)*



According to this taxonomy, learning motivation progresses along a continuum from *amotivation* to *intrinsic motivation*. *Amotivation* describes the complete absence of motivation — a lack of intention or willingness on the individual's part to act at all while *extrinsic motivation* refers to

performing an action for an instrumental purpose or a reward (Ryan & Deci, 2020). This type of motivation is further divided into four subcategories based on its level of internalization and its proximity to intrinsic motivation. The first of these, *external regulation*, is the most controlled type of motivation. In this state, the individual performs the action entirely to obtain external rewards or to avoid punishment. *Introjection*, another controlled type of motivation, is somewhat more internalized than the previous level: the individual acts to avoid guilt, anxiety, or shame, or to gain social approval. At the next level, *identification*, the individual acknowledges the value and importance of the action itself. At the final level, *integration*, this acknowledgment transforms into a complete assimilation of these values, bringing them into alignment with the individual's own identity. When an individual performs an action free of any external pressure — purely for their own enjoyment, interest, and personal satisfaction — this is *intrinsic motivation*, the most autonomous form of motivation. This SDT taxonomy by Ryan and Deci (2020) shows that an individual's success, willingness to engage, sustained commitment to an action, and long-term retention of what is learned are all directly related to how close their motivation lies to intrinsic motivation.

The Relationship Between Motivation and Learning with GenAI

Within the SDT framework, it can be maintained that sustainable academic success heavily relies on autonomous motivation. Therefore, it may be necessary to examine how emerging educational technologies, especially GenAI, a major breakthrough of recent years, relate to and influence learners' motivation. In fact, due to the inherent nature of GenAI technologies, there is a significant relationship between their use for high-quality, sustainable language education and motivation (Moybeka et al., 2023). Rather than a momentary effort, language learning requires sustained engagement, practice, and persistence from individuals—in other words, motivation (Ma & Chen, 2025). Therefore, integrating GenAI into language education is considered to be critical for sustaining that motivation. Viewed through the lens of SDT, GenAI-supported language education can meet three basic psychological needs that Ryan and Deci (2000) identify as indispensable for intrinsic motivation and arguably goes beyond them (Demirci et al., 2025).

For example, GenAI-supported language education directly addresses autonomy, one of these basic needs. When learning with GenAI, individuals can set their own pace, customize their learning, access personalized content, and study independently (Ma & Chen, 2025). However, SDT does not associate autonomy simply with working alone. According to Ryan

and Deci (2000), individuals must take ownership of and internalize the learning process as they make their own choices and learn freely. During this process, students must embrace GenAI as a learning partner and also be trained in AI literacy to eliminate risks such as potential technological dependency, the erosion of critical thinking skills, and the uncritical acceptance of inaccurate information generated by GenAI (Gulich & Chetveryk, 2025).

In terms of competence, GenAI's personalized feedback and error-correction functions can help language learners overcome their most common challenges in areas such as vocabulary, grammar, and pronunciation (Ma & Chen, 2025). Through learning from their mistakes in this way, students can gain a sense of competence in improving and learning the language (Moybeka et al., 2023), which in turn, enhances their levels of self-efficacy and intrinsic motivation.

Since GenAI systems create a neutral learning environment free of social judgment and emotional bias, they can help some students overcome language anxiety and the fear of being negatively evaluated for their mistakes (Gulich & Chetveryk, 2025). As a result, students can become less afraid to make mistakes and gain confidence in using the foreign language effectively. To prevent student social isolation, utilizing GenAI at this stage as a safe practice space—where concepts learned in class are put into practice — will most likely increase the classroom relatedness levels of students who have built self-confidence in such a safe environment (Ahmed et al., 2025).

The research on the relationship between motivation and GenAI shows in general that motivation plays a key role in the acceptance and effective use of GenAI in education (Wang et al., 2025). For instance, there is a direct relationship between the adoption and educational use of GenAI technologies, and increased intrinsic motivation (Ahmed & Sultan, 2026). Given the critical importance of this technology in education, the significance of motivation for this acceptance becomes evident.

There is also a substantial increase in the research specifically focusing on motivation and GenAI in the context of general education and foreign language learning. For instance, in an experimental study conducted with 60 EFL learners at the university level in China, Wei (2023) demonstrated that groups learning with GenAI support exhibited a significant increase in their second language motivation, academic achievement, and self-regulation strategies compared to groups learning through traditional methods. It was also found that the personalized learning opportunities and immediate feedback provided by GenAI attracted students' interest and strengthened their intrinsic motivation.

Similarly, in their study involving 49 pre-service English teachers in Poland, Kruk and Kałużna (2024) found that GenAI increased students' motivation and curiosity. GenAI, however, caused demotivation among some students. For example, since GenAI performs tasks so easily, it may reduce students' own engagement with those tasks and, for some students, create problems such as dependency on GenAI (Kruk & Kałużna, 2024).

Within the context of SDT and GenAI applications, Zhai and Nezakatgoo (2025), in a study involving 310 university students in China, demonstrated that using GenAI in English language teaching led to a substantial increase in students' intrinsic and autonomous motivation. According to their findings, while students enjoyed working with GenAI, this process increased their confidence in using the foreign language, decreased language anxiety, improved their language skills, and enhanced their cultural awareness of the target language (Zhai & Nezakatgoo, 2025).

In another study focusing on the use of GenAI and SDT in language teaching, Ma and Chen (2025) found in their research with 150 Chinese EFL university students that teacher scaffolding was of critical importance for the appropriate use of GenAI in language education. The results of this study revealed the experimental group that combined teacher support with GenAI managed to sustain their motivation longer than the group using GenAI alone (Ma & Chen, 2025).

In a scale-development study conducted in Türkiye, Demirci et al. (2025) investigated the motivation of university students learning English with GenAI within the framework of SDT's three basic factors — autonomy, competence, and relatedness. This study, which also established a valid and reliable instrument, found a strong correlation between students' level of GenAI adoption and their motivation (Demirci et al., 2025).

Expanding on the relationship between GenAI use and motivation, Zhen (2025) examined the role of teacher support alongside students' autonomous language-learning skills. That study demonstrated that while AI technology is a powerful catalyst for autonomy and competence, it cannot replace the human element in education. Hence, teacher support remained essential to satisfy students' psychological needs and ensure that AI technology effectively boosts internal motivation (Zhen, 2025).

Focusing specifically on technological acceptance, Wang et al. (2025) investigated the effects of Chinese EFL students' acceptance of GenAI on their motivation. Their study revealed that autonomous motivation, specifically the sub-dimensions of self-efficacy and self-satisfaction, served as a key mediator. Consequently, it was emphasized that proving technological acceptance alone is insufficient and must fulfill learners' internal

psychological needs to generate meaningful, persistent learning behavior (Wang et al., 2025).

Grounded in SDT, Shi and Zhang (2025) explored to what extend academic self-efficacy and resilience predict the learning motivation of Chinese university EFL students within AI-assisted educational environments. They found that the adaptive scaffolding, personalized feedback, and real-time analytics provided by AI tools fulfill learners' psychological need for competence, while psychological resilience supplies the stamina required to overcome technical problems and maintain long-term autonomous learning motivation.

Finally, Shi and Derakhshan (2026) analyzed the progress of GenAI use in language teaching from an SDT perspective over the course of a long teaching process. Their study revealed that the students' levels of autonomy, competence, and relatedness developed significantly over time due to AI integration. Their findings indicated that sustained interaction with adaptive AI technologies progressively satisfies learners' core psychological needs and systematically fosters long-term autonomous motivation throughout an instructional program.

While the existing literature provides valuable insights into the relationship between GenAI and motivation within the SDT framework, the majority of these studies have focused on EFL contexts. However, the motivation of pre-service teachers of German language — a population whose future professional role depends heavily on the long-term language proficiency — has not yet been fully examined within an SDT framework. Therefore, understanding the unique motivational profiles of pre-service German language teachers can provide insights into their learning situation as their professional context and linguistic needs may differ significantly from those of EFL learners.

Then, the present study aimed to answer the following research questions:

1. What are the pre-service German language teachers' motivation levels (intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation) regarding the use of GenAI tools?
2. Is there a relationship among autonomous motivation (intrinsic and identified), controlled motivation (introjected and external) and amotivation in the context of GenAI use by pre-service German language teachers?

Methodology

Research Design

To examine pre-service German language teachers' motivation toward GenAI technologies through the lens of SDT and to reveal the relationships among these types of motivation, this study employed a correlational survey design, one of the quantitative research methods. Survey design is a type of quantitative research that uses a scale to identify the attitudes, opinions, behaviors, or characteristics of a general population regarding a given topic or trend, through the sample to which the scale is administered (Creswell, 2012). Correlational survey design is used to draw inferences relevant to the research questions by investigating the relationships among variables within the survey (Fraenkel et al., 2011). Since this study used a pre-existing scale without modifying any of its items, this design was employed to examine the correlation among the sample's responses regarding the sub-dimensions of SDT (amotivation, extrinsic motivation, and intrinsic motivation).

Participants and Research Context

The participants of this study consisted of 94 students enrolled in the German Language Teaching Department at a state university in Türkiye. Of these students, 35 were male, 58 were female, and 1 preferred not to state their gender. The age range distributed mostly between 20 and 22 years old, with a mean age of around 21. There was one notable outlier at 48, which showed a mature or returning student in this cohort (see Figure 2). Regarding the year of study, first-year students formed the largest group with 33 participants, followed by second-year students, fourth-year students, and third-year students (see Table 1).

Figure 2: *Age Range of the Sample*

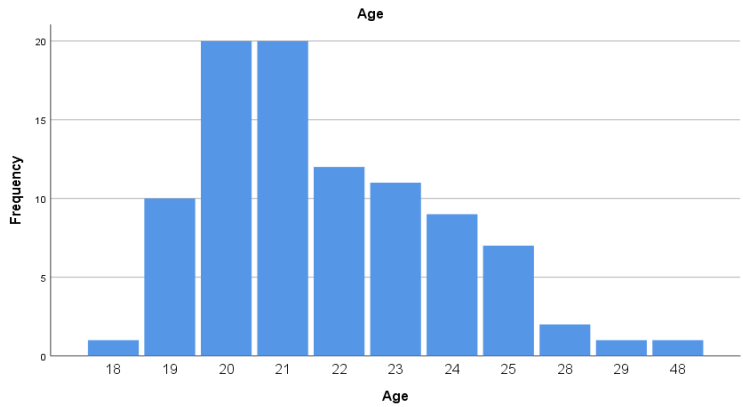


Table 1: *Study Year of German Language Teaching Students*

Year of study	f	%
First-year student	33	35.1
Second-year student	25	26.6
Third-year student	15	16.0
Fourth-year student	21	22.3
Total	94	100

Since the aim of this study was to gather data specifically from students in the German Language Teaching Department, only students from this department were selected, meaning that purposive sampling was employed (Patton, 2002).

Instrument and Data Collection

To collect the data, the *AI Motivation Scale (AIMS)* was utilized. This measurement tool was originally developed and validated by Li et al. (2025) with a large sample of over 1,000 university students in China (N = 1052). The scale has a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree) consisting of 20 items distributed across five subscales. Each subscale comprises four items developed in accordance with the SDT taxonomy to measure intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation. Sample items from the scale include: 'I enjoy learning with AI very much' (intrinsic motivation), 'Learning with AI could enhance my effectiveness' (identified regulation), and 'Learning with AI is meaningless' (amotivation).

The scale was administered in its original English form without any modifications. It was distributed and collected face-to-face in the classroom setting, in strict adherence to the principle of voluntary participation. The only additions to the survey were the questions regarding demographic information and a single open-ended question asking students to specify which GenAI tools they used (see Table 2). Notably, 75 out of 94 students — nearly 80% — reported using more than one tool. Only two students reported using no AI tool at all.

Table 2: *Students' AI Tools Usage (N = 94)*

Rank	AI Tool	n	%
1	ChatGPT	84	89.4
2	Gemini	77	81.9
3	NotebookLM	6	6.4
4	DeepSeek	5	5.3
4	DeepL	5	5.3
6	Grok	4	4.3
7	Copilot	3	3.2
8	ElevenLabs	2	2.1
8	Claude	2	2.1
8	None	2	2.1
11	Perplexity	1	1.1
11	Higgsfield	1	1.1
11	Banana Pro	1	1.1
11	Grammarly	1	1.1
11	Transkriptor	1	1.1
11	Gamma	1	1.1

Cronbach's alpha reliability coefficients were also calculated based on the data obtained from the current sample. Four of the five subscales demonstrated good to excellent internal consistency. Intrinsic motivation yielded a Cronbach's alpha of .929, identified regulation at .849, amotivation at .896, and external regulation at .799. The only subscale worth noting was introjected regulation at .677, which fell slightly below the conventional .70 threshold. However, since this is a 4-item scale, values in this range are considered acceptable for short subscales (Pallant, 2020).

Data Analysis

The data gathered were analyzed using the SPSS version 26. Descriptive statistics were computed to determine participants' levels of motivation toward learning with GenAI tools and to identify which SDT dimensions were relatively higher or lower. In this step, the mean and standard deviation values for each SDT subscale — intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation — were calculated to reveal students' levels of motivation across these

dimensions. Correlation analysis was then applied to determine the direction and strength of the relationships among these five motivational subscales of SDT.

Findings

This section presents the descriptive statistics and correlation analysis conducted on the data collected from 94 pre-service German language teachers through the AIMS.

Descriptive Statistics

To determine the motivation levels of pre-service German language teachers regarding GenAI use within the SDT framework, descriptive statistics were calculated. The values are presented in Table 3.

Table 3: *Motivation Profile of Pre-Service German Teachers (N = 94)*

Descriptive Statistics	M	SD
Intrinsic motivation	4.56	1.55
Identified regulation	4.44	1.44
Introjected regulation	1.90	.94
External regulation	1.97	1.14
Amotivation	2.51	1.50

The examination of the descriptive statistics revealed a meaningful pattern. Specifically, intrinsic motivation scored the highest at 4.57, and identified regulation followed closely at 4.45 — both above the midpoint of the scale, suggesting genuine positive engagement with AI learning. At the other end of the spectrum, introjected regulation was recorded at just 1.91, and external regulation at 1.98 — both well below the midpoint — while amotivation stood at 2.51. This indicated that pre-service German language teachers engaged with generative AI primarily because they found it interesting and recognized its value for their future, rather than because they felt pressured, guilty, or obligated to do so.

When an item-by-item basis was considered, certain items stood out prominently, either positively or negatively, both across the overall scale and within their respective subscales (see Appendix). For example, the highest-scoring item in the entire scale was Item 1 (‘I enjoy learning with AI very much’) with a mean of 4.86, while Item 5 (‘Learning with AI could enhance my effectiveness’) scored 4.84. These two items—one intrinsic and one identified—were essentially tied at the top. On the other end, the lowest-scoring item was Item 15 (‘I will get in trouble if I don’t

learn with AI') at just 1.56. This is an external regulation item concerning the fear of consequences, indicating a notable absence of such pressure among pre-service teachers. Aside from these, the item with the lowest mean score for introjected regulation was identified as Item 9 ('I will feel guilty if I don't learn with AI') at 1.77, and for amotivation, Item 20 ('Learning with AI is a waste of my time') at 2.16.

These results demonstrated consistency across both the subscale and item levels: interest and perceived value are driving AI engagement, not external pressure or guilt. Furthermore, students generally adopted GenAI in education and acknowledged that learning with AI was, on the whole, meaningful.

Correlation Analysis

Within the scope of this research study, a correlation analysis was also conducted to determine the direction and strength of the relationships among the motivational sub-dimensions of SDT. The statistical findings obtained from this analysis are summarized in Table 4.

Table 4: *Correlations Among the Motivation Types (N = 94)*

	Identified regulation	Introjected regulation	External regulation	Amotivation
Intrinsic motivation	.782**	.093	.209*	-.705**
Identified regulation		.069	.259*	-.660**
Introjected regulation			.539**	-.021
External regulation				-.301**

* $p < 0.05$
** $p < 0.01$

Table 4 displayed several significant relationships. First, intrinsic motivation and identified regulation correlated at .782, indicating a strong positive relationship. This means that students who enjoy learning with AI tend to find it personally valuable, showing that these two autonomous motivation types are closely linked.

Second, intrinsic motivation and amotivation correlated at -.705, representing a strong negative relationship. The more a student is genuinely interested in learning with AI, the less amotivated they tend to be, indicating that interest and disengagement do not coexist.

On the controlled motivation side, introjected and external regulation correlated at .539, reflecting a moderate positive relationship. This shows that students who feel guilty about not using AI also tend to feel externally pressured, indicating that these two controlled motivation types are related to one another. Notably, introjected regulation and amotivation showed essentially no relationship (r = -.021); that is, this non-significant correlation revealed that feeling guilty about AI use does not relate to a lack of intentionality.

Discussion

The aim of this study was to examine pre-service German language teachers' motivation for using GenAI in education within the framework of SDT. To this end, Li et al.'s (2025) AIMS was administered to collect data from German language teacher candidates. The findings revealed that pre-service teachers' levels of autonomous motivation (intrinsic motivation and identified regulation) were quite high whereas their levels of controlled motivation (introjected and external regulation) and amotivation were low. This general tendency indicated that GenAI was largely accepted by students as an educational tool without external pressure, and that using this tool met individuals' psychological needs. These findings are also consistent with the original AIMS validation study by Li et al. (2025). In their sample of over 1,000 Chinese university students, autonomous motivation types were also more prominent than controlled types (Li et al., 2025).

The strong relationship between intrinsic motivation and identified regulation showed that pre-service German language teachers who genuinely enjoy using AI recognize its practical value for their future classrooms, which implies that enjoyment and utility are not distinct motivators but rather reinforce each other. These findings align with those of Demirci et al. (2025) and Zhen (2025), who found high levels of autonomous motivation among students who adopted GenAI. The results of this study also showed that students used multiple GenAI tools simultaneously, which serves as an indicator of GenAI's acceptance among students. Similarly, Zhai and Nezakatgoo (2025) found that students who used GenAI for language learning with high motivation showed reduced language anxiety and improved language skills. Given that pre-service German language teachers' target language was German and that they would use this

language throughout their professional careers, they would require a more enduring type of motivation — one that, according to Ryan and Deci (2000), corresponds to intrinsic motivation and identified regulation. Since the results of this study revealed the higher means in intrinsic motivation and identified regulation, it can be inferred that there is a positive support of GenAI on language education.

GenAI offers features such as personalized and immediate feedback, individualized and goal-oriented learning, and the ability to set one's own pace, all of which support autonomous learning. Within the framework of SDT, these features provide advantages for students' autonomy and competence needs (Ma & Chen, 2025; Moybeka et al., 2023). For students to fully benefit from these advantages, further efforts are needed to increase their rates of GenAI acceptance (Ahmed & Sultan, 2026). For instance, to maintain students' interest, activities can be designed that are more engaging and action-oriented, and that prevent students from simply copying and pasting responses generated by GenAI in place of completing the assigned task themselves. Such activities would also reduce the risk of demotivation as highlighted by Kruk and Kałużna (2024). In addition, students can be encouraged to use GenAI as a kind of safe space for language learning. According to Ahmed et al. (2025), this allows students to learn from their mistakes and, through the confidence they gain, to develop a stronger sense of relatedness. Designing and offering suitable activities and strategies for that is acknowledged to be the responsibility of teachers (Ma & Chen, 2025).

On the contrary, the controlled motivation dimensions (introjected and external regulation) were found to be quite low in this study, and a positive correlation was found between them. The moderate correlation between external and introjected regulation signals that when academic requirements force students toward AI use, they tend to activate internal feelings of pressure or guilt. That is, these external and internal pressure mechanisms are intertwined. These results align with the findings of Li et al. (2025), as well. Nevertheless, controlled motivation types played a secondary role for students who already adopted GenAI in this study. As these technologies are further integrated into educational systems, this role is likely to diminish even more (Wei, 2023).

One of this study's key findings is the strong negative relationship between amotivation and intrinsic motivation. This shows that students who adopted GenAI and used it for educational purposes also found this use inherently meaningful. As Shi and Zhang (2025) similarly emphasize in their research, this positive form of motivation is likely to enhance students' levels of self-efficacy and resilience. However, in this context, it

can be inferred that *quality* of motivation matters, not just the *quantity*. The strong negative relationship between intrinsic motivation and amotivation indicated that cultivating genuine interest — not merely mandating AI use — is what keeps disengagement at minimum. Proper educational processes, training on AI literacy, and the aforementioned interactive activities might further improve students' motivational perceptions within the SDT framework over time (Shi & Derakhshan, 2026).

Finally, the type of motivation students hold should also be measured and monitored on an ongoing basis. For this purpose, in addition to the scale developed by Li et al. (2025), the scale developed by Demirci et al. (2025) can also be utilized. By focusing on different dimensions of SDT, both scales attempt to measure students' psychological needs as well as their motivation types. In this way, any shortcomings in achieving a sustained and lasting language education can be identified, and targeted solutions for these gaps can be developed over time. Although GenAI currently plays a critical role in education, as Ma and Chen (2025) emphasize in their study, teacher guidance is indispensable for both instruction and motivation. Intrinsic motivation can only be maintained robustly under the guidance and direction of teachers. Furthermore, *relatedness*, one of the three core needs of SDT, inherently encompasses the social relationship established with the teacher (Ryan & Deci, 2000). For this reason, using GenAI in language education calls for a hybrid model — one that does not marginalize the teacher in the educational process but instead brings with it a fundamental transformation in teacher roles.

Conclusion

This study examined the motivation of pre-service German language teachers regarding the use of GenAI technologies for educational purposes within the framework of SDT. The research results revealed that pre-service teachers generally used these technologies with high intrinsic motivation — without any external pressure, obligation, grade anxiety, or guilt — by recognizing their advantages, internalizing their value, and genuinely enjoying the process.

In light of these findings, it can be argued that the right type of motivation plays a significant role in ensuring sustained and lasting engagement with GenAI in education. To cultivate this type of motivation, teachers must adapt to their new roles within the changing educational landscape. It is apparent that learning solely through GenAI is insufficient. To achieve lasting educational outcomes, particularly in human-centric disciplines like language teaching, the teacher's role may be more vital than ever before. Furthermore, teacher education curricula should not only focus on

the mechanics of GenAI usage; rather, they should support pedagogical shifts designed to strengthen autonomous motivation. Therefore, students, teachers, and faculties must work collaboratively to ensure the quality of this mode of education, developing policies directed toward AI literacy, human-AI collaboration, and action-oriented GenAI applications in teaching to sustain higher motivation levels in different dimensions.

In conclusion, it should be note that this study had some limitations. First, it was a quantitative study conducted with a limited sample size. Future studies supplementing these quantitative data with qualitative interviews could provide deeper insights, shifting the focus from 'to what extent' students exhibit this type of motivation to 'why' they experience it. Similarly, future research could examine how in-service German language teachers—rather than just pre-service teachers—utilize GenAI in their instruction, for what purposes, and how this impacts motivation within the SDT framework.

Declaration on the use of artificial intelligence

The author acknowledges the use of DeepL and Gemini but only for checking grammatical accuracy and clarity. All the content of this paper represents the original work of the authors.

References

Ahmed, I., Ghafoor, A., Liliuara, D., & Akyuningrum, V. Q. (2025). The impact of AI learning tools on ESL learners' motivation and success: A systematic literature review. *English Learning Innovation (englie)*, *6*(1), 134–142. https://doi.org/10.22219/englie.v6i1.38961

Ahmed, A., & Sultan, A. (2026). Examining the impact of generative AI on student motivation and engagement: The mediating role of autonomy-support and autonomous motivation in education. *Frontiers in Psychology*, 17, 1852265. https://doi.org/10.3389/fpsyg.2026.1852265

Aly, A. H., Setiadi, S., & Zuriyati, Z. (2025). Exploring AI's ethical impact on teacher-student dynamics in English language learning. *Journal La Edusci*, *6*(1), 1–16. https://doi.org/10.37899/journallaedusci.v6i1.1703

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, *32*(7), 3240–3256. https://doi.org/10.1080/10494820.2023.2172044

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2024). Integration of generative artificial intelligence in higher education: Best practices. *Education Sciences, 15*(1), 32. https://doi.org/10.3390/educsci15010032

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education*, *1*(3), 158–167. https://doi.org/10.47852/bonviewIJCE42022495

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. United States of America: Pearson Education.

Demirci, H., Kara, M., & Korkmaz, Ö. (2026). Measuring motivation for generative AI-based language learning: Scale development and validation based on the self-determination theory. *International Journal of Applied Linguistics*, *36*, 1425–1436. https://doi.org/10.1111/ijal.12868

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, *78*(3), 273–284.

Espinar Redondo, R., & Ortega Martín, J. L. (2015). Motivation: The road to successful learning. *PROFILE Issues in Teachers' Professional Development*, *17*(2), 125–136. http://dx.doi.org/10.15446/profile.v17n2.50563

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*, 489–530. https://doi.org/10.1111/bjet.13544

Fraenkel, J. R. et al. (2016). *How to design and evaluate research in education*. New York: McGraw-Hill.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House Publishers.

Gulich, O., & Chetveryk, V. (2025). Generative artificial intelligence as a tool for psychological and motivational support in foreign language learning. *Educational Challenges*, *30*(2), 248–261. https://doi.org/10.34142/2709-7986.2025.30.2.19

Kruk, M., & Kałużna, A. (2025). Investigating the role of AI tools in enhancing translation skills, emotional experiences, and motivation in L2 learning. *European Journal of Education*, *60*, Madde e12859. https://doi.org/10.1111/ejed.12859

Li, J., King, R. B., Chai, C. S., Lee, V. W. Y., & Li, R. (2026). AI and student motivation: A scoping review. *Journal of Research on Technology in Education*. Advance online publication. https://doi.org/10.1080/15391523.2026.2672515

Li, J., King, R. B., Chai, C. S., Zhai, X., & Lee, V. W. Y. (2025). The AI Motivation Scale (AIMS): A self-determination theory perspective. *Journal of Research on Technology in Education*. Advance online publication. https://doi.org/10.1080/15391523.2025.2478424

Lukianenko, N. (2025). Motivational pathways in language learning: Psychological perspective. *Вісник ХНАДУ*, *108*, 318–324. https://doi.org/10.30977/BUL.2219-5548.2025.108.0.318

Ma, Y., & Chen, M. (2025). The human touch in AI: Optimizing language learning through self-determination theory and teacher scaffolding. *Frontiers in Psychology*, *16*, Madde 1568239. https://doi.org/10.3389/fpsyg.2025.1568239

Moybeka, A. M. S., Syariatin, N., Tatipang, D. P., Mushthoza, D. A., Dewi, N. P. J. L., & Tineh, S. (2023). Artificial intelligence and English classroom: The implications of AI toward EFL students' motivation. *Edumaspul: Jurnal Pendidikan*, *7*(2), 2444–2454.

Nadim, M. A., & Di Fuccio, R. (2025). Unveiling the potential: Artificial intelligence's negative impact on teaching and Research considering ethics in higher education. *European Journal of Education, 60*(1). https://doi.org/10.1111/ejed.12929

Nelson, A. S., Santamaría, P. V., Javens, J. S., & Ricaurte, M. (2025). Students' perceptions of Generative Artificial intelligence (GENAI) use in academic writing in English as a foreign language. *Education Sciences*, *15*(5), 611. https://doi.org/10.3390/educsci15050611

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*(7th ed.). McGraw-Hill Education.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage Publications.

Riaz, S., Ayaz, N., & Khan, M. U. (2026). Artificial intelligence and self-regulated learning: Implications for student success in contemporary higher education. In *AI and the future of learning - Psychological perspectives on student success and teacher resilience*. IntechOpen. https://doi.org/10.5772/intechopen.1016671

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, Madde 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Shi, L., & Derakhshan, A. (2026). Exploring the role of AI in EFL students' autonomy, relatedness, and competence through the lens of self-determination theory (SDT): A latent growth curve modeling (LGCM) analysis. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70272

Shi, S., & Zhang, H. (2025). EFL students' motivation predicted by their self-efficacy and resilience in artificial intelligence (AI)-based context: From a self-determination theory perspective. *Learning and Motivation*, *91*, 102151. https://doi.org/10.1016/j.lmot.2025.102151

Wang, Q., Amini, M., & Fu, Z. (2025). AI acceptance and Chinese EFL learners' behavioral engagement with mediating effects of motivation. *Scientific Reports*, *15*, Article 33310. https://doi.org/10.1038/s41598-025-11305-2

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, Madde 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Zhai, Y., & Nezakatgoo, B. (2025). Evaluating AI-powered applications for enhancing undergraduate students' metacognitive strategies, self-determined motivation, and social learning in English language education. *Scientific Reports*, *15*, 35199. https://doi.org/10.1038/s41598-025-19118-z

Zhen, X. (2025). Application of artificial intelligence in English learning: The interactive effects of teacher support, learning motivation, and autonomous learning ability. *Acta Psychologica*, *261*, 105960. https://doi.org/10.1016/j.actpsy.2025.105960

Appendix

Descriptive Statistics of Items

Items	N	M	SD
1. I enjoy learning with AI very much.	94	4.86	1.66
2. Learning with AI is fun for me.	94	4.79	1.67
3. I feel satisfied through learning with AI.	94	4.32	1.71
4. Exploring knowledge with AI is fascinating to me.	94	4.30	1.78
5. Learning with AI could enhance my effectiveness.	94	4.84	1.62
6. Learning with AI is important for my academic performance.	94	4.54	1.74
7. Learning with AI is important for my future career.	94	4.06	1.77
8. Learning with AI is important in helping me deal with future challenges.	94	4.35	1.81
9. I will feel guilty if I don't learn with AI.	94	1.77	1.21
10. I feel a sense of obligation to learn with AI because others expect it of me.	94	2.22	1.44
11. I will feel ashamed if I cannot learn with AI.	94	1.86	1.38
12. I will feel ashamed if others think I am incapable of learning with AI.	94	1.79	1.25
13. I have no other choice but to learn with AI.	94	1.72	1.33
14. Learning with AI is one of the rules.	94	2.17	1.60

15. I will get in trouble if I don't learn with AI.	94	1.56	1.02
16. I'm required to learn with AI.	94	2.45	1.73
17. Learning with AI is meaningless.	94	2.37	1.60
18. I cannot see why I need to learn with AI.	94	2.80	1.83
19. I don't know what I am getting out of learning with AI.	94	2.71	1.76
20. Learning with AI is a waste of my time.	94	2.16	1.70
Valid N (listwise)	94		

AI as Support, Not Substitute: English Instructors' Perceptions of ChatGPT in Foreign Language Teaching

Şule Çınar Yağcı [1] & Tuğba Aydın Yıldız [2]

Abstract

This qualitative case study explored how English language instructors make sense of ChatGPT and other artificial intelligence (AI) applications in foreign language teaching following a four-hour in-service professional development program. Four instructors at the School of Foreign Languages of a state university in Türkiye participated in semi-structured interviews, which were analyzed using thematic content analysis. The findings revealed a simultaneously enabling and cautionary view of AI. Participants valued AI for facilitating material development, expanding access to instructional resources, reducing preparation time, supporting language learning, and promoting learner autonomy. However, these perceived benefits were closely intertwined with concerns about privacy, academic integrity, originality, student dependency, and reduced cognitive engagement. Participants also viewed AI as reshaping rather than replacing teachers' professional roles, increasing the importance of pedagogical judgment, critical evaluation, and AI literacy. Practice-based professional development and clear institutional guidance were considered necessary for responsible implementation. Overall, the study shows that the educational value of AI depends less on technological affordances alone than on how they are mediated through teacher expertise, learner responsibility, ethical safeguards, and institutional support.

Keywords: ChatGPT, foreign language teaching, teacher perceptions, AI literacy, professional development.

Introduction

Artificial intelligence (AI) applications have recently been addressed in education not merely as technical innovations, but also as tools that support teaching and learning processes in meaningful ways. These applications provide students with opportunities for personalized learning, while helping teachers and researchers offer rapid feedback, support assessment processes, and create more interactive learning environments. In this

[1] Lecturer Dr., Zonguldak Bülent Ecevit University, School of Foreign Languages, Department of Foreign Languages, French Language Program, ORCID: 0000-0003-4456-3174, *sule.yagci@beun.edu.tr*, Türkiye

[2] Asst. Prof. Dr., Zonguldak Bülent Ecevit University, Ereğli Faculty of Education, Department of Foreign Language Education, English Language Teaching Program, ORCID: 0000-0001-5248-2484, *tugba.ayildiz@beun.edu.tr*, Türkiye

regard, AI technologies may provide learning pathways adapted to students' individual needs with the aim of improving their learning pace and academic achievement (Çınar Yağcı & Genç, 2024, p. 86). For this reason, AI has become increasingly visible both in classroom practices and in out-of-class learning processes. Generative AI tools, in particular, have attracted attention in foreign language teaching because they can produce texts, provide feedback, organize content, and support learning according to individual needs. In skill-based language teaching, these tools further enrich the learning process by enabling individualized content production, offering immediate feedback, encouraging metacognitive strategies, and supporting learner autonomy (Sarı, 2025, p. 246). However, the educational value of these tools is not limited to the speed and convenience they offer. Rather, the use of AI constitutes a multidimensional field that directly affects the role of the teacher, the learner's responsibility in the learning process, the understanding of academic integrity, and institutional decisions on this issue.

Within this transformation, ChatGPT holds particular significance for foreign language teaching and learning. This is because the tool can generate natural language, revise texts, interact through question-and-answer exchanges, and provide students with immediate support. Through these features, ChatGPT offers various areas of use in the language learning process. For this reason, ChatGPT is considered in foreign language classrooms not merely as a technical tool, but as a new learning partner that encourages a reconsideration of teaching and learning processes (Çınar Yağcı, 2026). Similarly, it has been stated that AI-supported personalized learning helps students manage their own learning processes more autonomously and strengthens their motivation (Özçelik & Çınar Yağcı, 2024). From this perspective, ChatGPT can be viewed as a supportive tool in students' processes of practice, self-paced learning, receiving feedback, and experimenting with different forms of language use. However, this support should not mean that students transfer their learning responsibility entirely to AI. Therefore, the role of ChatGPT in foreign language teaching should be addressed not only in terms of learner autonomy, but also in relation to preserving students' effort to think and their ability to make critical judgments.

Recent studies show that AI tools are being used increasingly widely by foreign language learners. Bayraktar Balkır and Zehir Topkaya (2025) examined English language learners' habits of using AI, their perceptions of these tools, and their educational needs in the Turkish context, and revealed that AI has begun to occupy a more visible place in students' language learning practices. This finding indicates that students are no longer limited to the resources provided by the teacher, but also benefit from AI-

supported tools in their learning processes. Similarly, Dizon (2026) stated that ChatGPT can support self-directed foreign language learning and showed that this tool may contribute to students' ability to sustain their own learning processes. When these two studies are considered together, it is understood that ChatGPT offers important support for learner autonomy and individual study processes. Nevertheless, the question of whether students use AI as a conscious learning tool or regard it as a shortcut that reduces their learning responsibility remains important.

The contribution of ChatGPT to foreign language learning has become particularly evident in the context of writing skills. This tool can produce grammatically accurate sentences and contextually appropriate texts, provide feedback on students' written products, and offer suggestions for improving the quality of writing (Chen, 2023; Özçelik & Çınar Yağcı, 2024; Üstünbaş, 2024). In this respect, ChatGPT may contribute not only to students' text production, but also to their ability to review and improve the texts they write. Indeed, Ataş, Cengiz, and Çelik (2024) examined the extent to which ChatGPT generates feedback on content and organization components in English comparison-and-contrast essays, and showed that this tool can provide feedback capable of complementing teacher support in the writing process. However, the study also highlights the importance of teacher supervision. Similarly, Barrot (2023) emphasizes that ChatGPT has strong potential to support second language writing, but that it should be used carefully in terms of originality, academic integrity, and students' critical evaluation skills. Shi, Chai, Zhou, and Aubrey (2026), by comparing ChatGPT with automated writing evaluation tools, examined the role of this tool in students' writing development and their self-perception as second language writers. When these studies are considered together, it becomes clear that ChatGPT can be used as an effective feedback and support tool in writing skills; however, this use should be addressed together with teacher guidance and ethical principles.

The literature includes experimental studies examining the effects of ChatGPT on writing, grammar, reading, and vocabulary learning; studies comparing teacher feedback with ChatGPT-generated feedback; and research focusing on learners' attitudes, experiences, and intentions to use this tool (Guo & Wang, 2024; Özdemir Çiftçi & Tilbe, 2025; Werdiningsih et al., 2024; Xiao & Zhi, 2023). These studies show that ChatGPT has been evaluated in foreign language learning across different skill areas and learning dimensions. A similar situation is also observed in relation to speaking skills. Yıldız (2024) examined the effect of ChatGPT use on English language learners' speaking self-efficacy and revealed that this tool may provide a learning environment that supports students' confidence in speaking. This finding is important in terms of enabling students

to engage in more speaking practice outside the classroom, reduce their anxiety about making mistakes, and gain more frequent exposure to the target language. Karataş et al. (2024), in their qualitative case study conducted in the Turkish context, also showed that ChatGPT contributes to the foreign language learning process in terms of writing, grammar, and vocabulary learning, while also bringing certain limitations and challenges. Therefore, the literature indicates that ChatGPT can support different language skills. However, the conditions under which this support becomes more effective, reliable, and meaningful for teaching remain an issue that still needs to be discussed.

From the perspective of teachers, ChatGPT stands out as an important support tool in the processes of lesson preparation and classroom activity design. This tool can assist teachers in many instructional processes, such as preparing lesson plans, activities, dialogues, role-play tasks, and vocabulary and grammar exercises (Büyüksaraç, 2026). Studies based on teachers' views also support this contribution. Orhan and Cinar Yagci (2026) state that ChatGPT may make meaningful contributions to foreign language teaching by saving time, supporting learner autonomy, strengthening self-regulated learning, increasing motivation and self-confidence, offering personalized learning experiences, sustaining out-of-class language practice, and developing cultural awareness. These findings show that ChatGPT is not merely a tool that accelerates material production for teachers. In addition, ChatGPT also contributes to teachers' ability to prepare more varied and flexible content suited to students' levels, needs, and learning goals.

However, these opportunities offered by ChatGPT do not diminish the role of the teacher. On the contrary, they make the teacher's responsibilities for selection, revision, evaluation, and guidance more visible. Öztürk Demir and Çetin (2025) examined instructors' views on the use of ChatGPT in English language teaching and emphasized that this tool should be used under teacher guidance, within ethical boundaries, and in a conscious manner. This finding shows that the teacher is not placed in a passive position in relation to AI; rather, the teacher assumes a more selective, evaluative, and guiding role. The qualitative studies conducted by Gao et al. (2024) and Orhan and Cinar Yagci (2026) also examine how teachers make sense of ChatGPT in classroom practices, what benefits and limitations they experience, and how they evaluate this tool in terms of ethical, reliable, and instructional use. These studies are considered important for understanding the changing professional role of the teacher in the age of AI.

The opportunities offered by ChatGPT for material development and the production of cultural content are also important for foreign language teaching. This tool can provide teachers with an initial draft in the process of preparing cultural content and instructional materials; in this respect, it is regarded as a supportive resource that promotes cultural awareness and increases content variety (Yalçın & Çınar Yağcı, 2024). However, content generated by AI should not be used directly in the classroom. Such content needs to be carefully reviewed by the teacher, revised where necessary, and adapted to the classroom context (Al-Maawali, 2026; Çınar Yağcı & Aydın Yıldız, 2023). This is because materials used in foreign language teaching should be evaluated not only in terms of linguistic accuracy, but also in relation to students' proficiency level, cultural appropriateness, the aims of the lesson, and the characteristics of the class. Therefore, a text, activity, or dialogue prepared by ChatGPT may serve as a useful starting point for the teacher; however, the final decision should always be based on the teacher's professional knowledge and evaluation.

At this point, the limitations and risks of AI tools should be carefully considered. The literature indicates that ChatGPT may sometimes produce inaccurate, incomplete, decontextualized, or unverified information (Çınar Yağcı & Aydın Yıldız; Orhan & Cinar Yagci, 2026). Findings based on teachers' views also support this concern. Accordingly, ChatGPT may, in some cases, provide incorrect information, fabricated references, grammatically problematic sentences, or responses that are not appropriate to the given context (Orhan & Cinar Yagci, 2026). This may create important risks, particularly in relation to academic writing, source use, assessment and evaluation, cultural content production, and student assignments. Therefore, the use of ChatGPT in foreign language teaching should not be evaluated only in terms of the types of content the tool can generate. How the generated content is checked, revised, and used within ethical and instructional principles should also be taken into account to the same extent.

The international literature also shows that ChatGPT should be addressed in a balanced manner in foreign language teaching. Yang and Li (2024) systematically reviewed studies on the use of ChatGPT in second language learning and revealed that this tool can be used in interaction, feedback, instructional support, and autonomous learning processes. This finding indicates that the function of ChatGPT in foreign language teaching is not limited to a single skill or area of use. In addition, Kır and Tekin (2025) examined the main features of AI applications used in foreign language teaching and emphasized that these tools should be evaluated in terms of their purpose, areas of use, and contribution to the learning process. When these two studies are considered together, it becomes clear

that for AI tools to be useful in foreign language teaching, teachers need to become familiar with these tools, critically evaluate the outputs they generate, and use them in ways that are appropriate to the classroom context.

When the studies in the literature are considered together, AI appears to be both a powerful support tool in foreign language teaching and a field of change that needs to be managed carefully. On the one hand, ChatGPT offers students opportunities for personalized study, rapid feedback, continued learning beyond the classroom, and the development of different language skills. On the other hand, it saves teachers time in lesson planning, material preparation, designing role-play activities, developing cultural content, and supporting student work. However, these opportunities also bring several important challenges. Academic integrity, originality, data security, the production of inaccurate information, the use of fabricated references, contextual appropriateness, and students' excessive dependence on AI are among the most prominent concerns. Therefore, examining instructors' perceptions of ChatGPT only through positive or negative attitudes is not sufficient. A more comprehensive evaluation requires studies that address instructors' experiences, concerns, changing professional roles, and support needs together.

Although previous studies have examined learners' use of ChatGPT, writing feedback, autonomy, and skill-based outcomes, fewer studies have explored how English instructors interpret ChatGPT after a structured professional development experience, particularly in relation to teacher role transformation, ethical concerns, and institutional support needs in Turkish higher education.

Based on this need, the present study examines the perceptions, experiences, concerns, and professional development needs of four English instructors working at the School of Foreign Languages of a state university in Türkiye regarding the use of ChatGPT and AI applications in foreign language teaching.

Significance of the Study

This study is significant in that it reveals how ChatGPT and AI applications are perceived by instructors in the context of foreign language teaching. The study examines in detail which aspects of these tools instructors consider useful and which issues raise concerns for them. The literature includes various studies examining students' habits of using AI, the effects of ChatGPT on writing and speaking skills, learner autonomy, and feedback processes. However, there is still a need for qualitative studies

that address English instructors' perceptions of ChatGPT after an in-service training program, their experiences, concerns, views on the changing role of the teacher, and expectations for institutional support together. Therefore, this study focuses on this gap in the literature and aims to understand instructors' views on AI directly through their own experiences.

At this point, the significance of the study is not limited to identifying instructors' positive or negative views. Rather than approaching the use of AI from a one-sided, benefit-oriented perspective, the study addresses it together with its educational, ethical, and professional dimensions. It is known that ChatGPT offers important opportunities, such as saving time, facilitating material preparation, supporting language learning processes, and contributing to learner autonomy. However, for these opportunities to be used effectively and reliably in educational settings, the possible risks also need to be examined carefully. Therefore, issues such as academic integrity, originality, data security, students' excessive dependence on AI, the production of inaccurate information, and teacher supervision are also included within the scope of the study. In this way, the study evaluates the place of ChatGPT in foreign language teaching not only in terms of its benefits, but also with regard to responsible use and the quality of the teaching process.

This evaluation is also important for understanding the changing professional role of the teacher. Tools such as ChatGPT should not be regarded as systems that replace teachers. On the contrary, these tools should be considered as forms of support that make the teacher's responsibilities for selection, revision, evaluation, and guidance more visible. In this context, the study reveals instructors' need for AI literacy, critical evaluation skills, and practice-based professional development activities. The findings are expected to contribute to the more functional design of in-service training programs on the use of AI in foreign language teaching. In addition, the findings of the study are expected to guide the development of clear institutional principles for AI use and to support students in using AI more consciously and responsibly.

Research Questions

This study sought to answer the following research questions:

1. How do English instructors evaluate the contributions of ChatGPT and AI applications to foreign language teaching?
2. What difficulties, risks, and concerns do English instructors express regarding the use of ChatGPT and AI?

3. How do English instructors make sense of the changing professional role of the teacher in the age of AI?
4. What are English instructors' professional development and institutional support needs regarding the use of ChatGPT and AI applications?

Methodology

Research Design

This study adopted a qualitative case study design to explore English language instructors' perceptions, experiences, and concerns regarding the integration of AI into language teaching following a professional development program. A qualitative case study is particularly suitable for examining a contemporary phenomenon within its real-life context and for gaining an in-depth understanding of participants' experiences (Creswell & Poth, 2018). In the present study, the case was bounded by a four-hour in-service training program on ChatGPT and AI applications in foreign language education and the experiences of the instructors who participated in this program.

The study aimed to explore how instructors perceived the pedagogical benefits, challenges, ethical implications, and professional demands associated with AI-supported language teaching. Given the exploratory nature of these objectives, a qualitative approach was deemed appropriate for capturing participants' subjective interpretations and experiences in depth (Creswell & Poth, 2018).

Participants

The participants of the study consisted of four English language instructors employed at the School of Foreign Languages of a state university in Türkiye. All participants voluntarily attended a four-hour professional development workshop entitled "ChatGPT and Artificial Intelligence in Language Education" and subsequently agreed to participate in semi-structured interviews.

The participants ranged in age from 39 to 44 years (M = 41.75). Two participants were female and two were male. Three participants held a master's degree, while one participant held a doctoral degree. All participants had extensive experience in English language teaching and were actively involved in tertiary-level language education. To ensure anonymity and confidentiality, participants were assigned pseudonyms (P1–P4) throughout the study.

Data Collection Instrument

Data were collected through a semi-structured interview form developed by the researchers. Semi-structured interviews are widely used in qualitative research because they allow participants to express their experiences and perspectives in detail while maintaining a common structure across interviews (Creswell & Poth, 2018).

The interview protocol consisted of open-ended questions designed to investigate instructors' perceptions of AI integration in English language teaching. The questions focused on four primary dimensions: (a) perceived pedagogical benefits of AI, (b) ease of adoption and implementation, (c) concerns regarding ethics, privacy, dependency, and academic integrity, and (d) evaluations of the AI training program. The semi-structured format enabled participants to elaborate on issues they considered important while ensuring consistency in data collection.

Data Collection Procedure

The data were collected following the completion of a four-hour in-service training program on ChatGPT and AI applications in foreign language education. The training introduced participants to a variety of generative AI tools and demonstrated their potential applications in language teaching, including lesson planning, material development, assessment, feedback provision, and classroom activities.

Upon completion of the training, four volunteer instructors participated in individual semi-structured interviews. The interviews were conducted in Turkish to facilitate richer and more authentic responses. All interviews were audio-recorded with participants' permission and subsequently transcribed verbatim. The resulting transcripts were carefully reviewed and checked for accuracy before being imported into MAXQDA 2020 for analysis.

Data Analysis

The interview data were analyzed using thematic content analysis supported by MAXQDA 2020 software. Thematic analysis is a flexible and systematic approach that enables researchers to identify, organize, analyze, and report recurring patterns within qualitative data (Braun & Clarke, 2006). The analysis followed the six-phase procedure proposed by Braun and Clarke (2006). First, the researchers familiarized themselves with the data through repeated reading of the interview transcripts. Second, initial codes were generated by identifying meaningful units of text related to participants' experiences and perceptions of AI-supported

language teaching. Third, similar codes were grouped into broader categories and potential themes. Fourth, the preliminary themes were reviewed and refined to ensure internal coherence and meaningful distinctions between themes. Fifth, the themes were defined and named to accurately represent the underlying patterns within the dataset. Finally, the themes were interpreted and supported with representative quotations from participants.

Throughout the analysis process, MAXQDA facilitated the systematic organization, retrieval, comparison, and visualization of coded segments. The software also supported the examination of code frequencies and thematic relationships, thereby enhancing the rigor and transparency of the analytical process. The coding process was informed by recommendations for qualitative coding proposed by Saldaña (2021). Similar codes were continuously compared and refined through an iterative process until meaningful thematic structures emerged. In addition, the procedures suggested by Miles, Huberman, and Saldaña (2020) regarding data reduction, data display, and conclusion drawing guided the overall analytical process.

Trustworthiness

Several strategies were employed to enhance the trustworthiness of the study. Following the criteria proposed by Creswell and Poth (2018), credibility was strengthened through prolonged engagement with the data and repeated examination of interview transcripts. All interviews were transcribed verbatim to preserve participants' original meanings and perspectives.

Dependability was supported through the systematic coding procedures implemented in MAXQDA and the maintenance of detailed coding records. Confirmability was enhanced through continuous reflection during the coding and theme development processes. Finally, representative quotations were included in the findings section to provide evidence for interpretations and allow readers to evaluate the connection between the raw data and the reported themes.

Ethical Considerations

Participation in the study was entirely voluntary. Prior to the interviews, participants were informed about the purpose of the study and provided informed consent. Anonymity and confidentiality were ensured through the use of participant codes (P1–P4). All interview recordings and transcripts were stored securely and used solely for research purposes. Also,

the ethical approval for this study was obtained from the Human Research Ethics Committee of Zonguldak Bülent Ecevit University on November 1, 2024 (Protocol No. 842).

Findings

The analysis of the interview data resulted in four overarching themes: Perceived Pedagogical Benefits of Artificial Intelligence, Challenges and Risks Associated with Artificial Intelligence, Teacher Adaptation and Changing Professional Roles, and Professional Development and Institutional Support Needs. Figure 1 presents the hierarchical coding model generated through thematic content analysis in MAXQDA. To provide a clearer understanding of these themes and the codes that constitute them, representative excerpts from participants' views were included throughout the findings section. The use of direct quotations contributed to a more detailed representation of participants' experiences and strengthened the presentation of the findings as grounded in the original interview data.

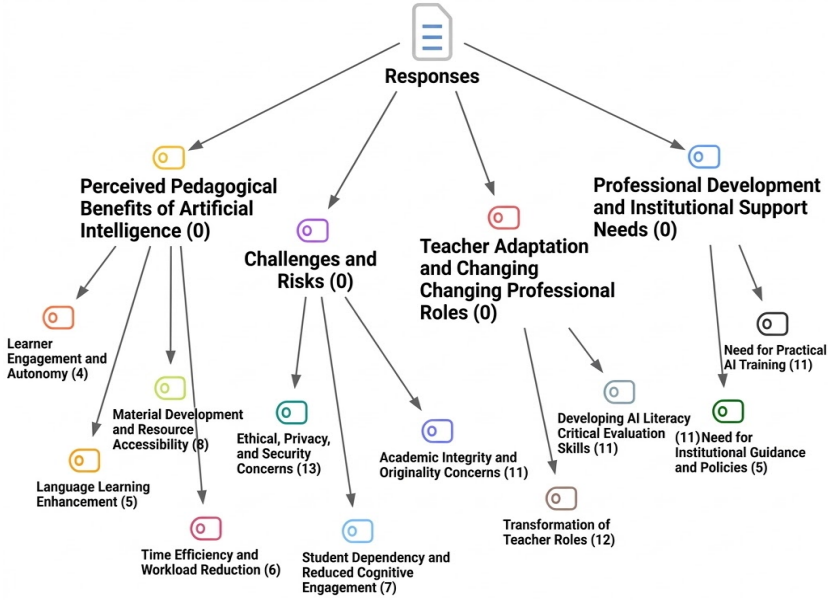


Figure 1. Hierarchical Coding-Thematic Model

Figure 1 presents the hierarchical coding–thematic model derived from participants' responses regarding the integration of AI into English language teaching. The model includes four overarching themes and eleven related sub-codes. Together, these themes reflect participants' perceptions of AI's pedagogical value, the risks they associate with its use, the changing professional role of teachers, and the support needed for responsible AI integration.

The first theme, Perceived Pedagogical Benefits of Artificial Intelligence, includes four sub-codes: Material Development and Resource Accessibility (f=8), Time Efficiency and Workload Reduction (f=6), Language Learning Enhancement (f=5), and Learner Engagement and Autonomy (f =4). Among these, Material Development and Resource Accessibility was the most frequently referenced code. This indicates that participants regarded AI as a useful resource for preparing instructional materials and accessing educational content more easily. Time Efficiency and Workload Reduction was also emphasized, suggesting that instructors perceived AI as a tool that can simplify routine teaching tasks and reduce preparation time. Participants further noted that AI may support language learning processes, increase learner engagement, and encourage learner autonomy. Overall, this theme shows that participants associated AI with several pedagogical opportunities, particularly when its outputs are reviewed and adapted by teachers.

The second theme, Challenges and Risks Associated with Artificial Intelligence, consists of three sub-codes: Ethical, Privacy, and Security Concerns (f=13), Academic Integrity and Originality Concerns (f =11), and Student Dependency and Reduced Cognitive Engagement (f=7). Ethical, Privacy, and Security Concerns emerged as the most frequently cited code within this theme. This finding suggests that participants were particularly concerned about data protection, responsible AI use, and the possible misuse of personal information. Academic Integrity and Originality Concerns were also strongly emphasized. Participants expressed worries about plagiarism, authorship, and the difficulty of distinguishing between student-generated and AI-generated work. In addition, they noted that excessive reliance on AI may weaken students' independent thinking and reduce their cognitive engagement. These findings indicate that AI integration requires careful attention to ethical, academic, and pedagogical safeguards.

The third theme, Teacher Adaptation and Changing Professional Roles, includes two sub-codes: Transformation of Teacher Roles (f=12) and Developing AI Literacy and Critical Evaluation Skills (f=11). Participants frequently stated that AI is reshaping traditional understandings of

teaching. They viewed the teacher not only as a transmitter of knowledge, but also as a facilitator, guide, and evaluator of AI-generated content. Closely related to this transformation was the need to develop AI literacy and critical evaluation skills. Participants emphasized that teachers should be able to assess the accuracy, relevance, and pedagogical appropriateness of AI-generated outputs. They also highlighted the importance of guiding students in the responsible use of AI tools. This theme shows that AI integration is not merely a technical issue; it also requires new professional competencies and a clear redefinition of teachers' roles.

The fourth theme, Professional Development and Institutional Support Needs, comprises two sub-codes: Need for Practical AI Training (f=11) and Need for Institutional Guidance and Policies (f=5). The most frequently referenced code in this theme was the Need for Practical AI Training. This finding indicates that participants expected hands-on professional development activities focused on classroom-based AI use. They did not view theoretical knowledge alone as sufficient. Participants also emphasized the importance of institutional guidance, ethical frameworks, and clear policy regulations. These expectations suggest that responsible AI integration should be supported not only by individual teacher competence, but also by institutional structures and shared guidelines.

Overall, the thematic structure demonstrates that instructors' perceptions of AI in English language teaching are multidimensional. Participants recognized AI's potential to support material development, save time, enhance language learning, and foster learner autonomy. At the same time, they expressed concerns about ethics, privacy, academic integrity, student dependency, and changing professional responsibilities. The findings therefore suggest that effective AI integration requires a balanced approach. AI should be used as a supportive tool, not as a substitute for teacher expertise. Its meaningful use depends on continuous professional learning, critical AI literacy, ethical awareness, and clear institutional policies.

Theme 1: Perceived Pedagogical Benefits of Artificial Intelligence

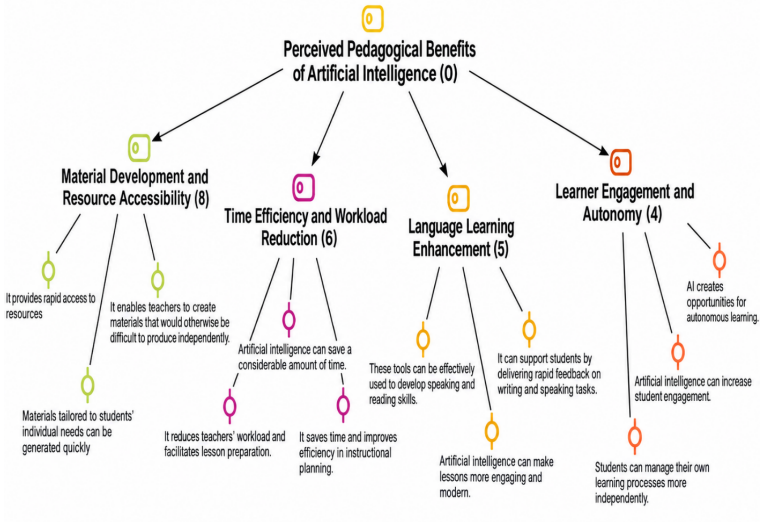


Figure 2. Theme 1: Perceived Pedagogical Benefits of Artificial Intelligence (by MAXQDA)

As shown in Figure 2, the theme of Perceived Pedagogical Benefits of Artificial Intelligence reflects participants' views on the educational value of AI in English language teaching. Four sub-codes emerged under this theme: Material Development and Resource Accessibility (f=8), Time Efficiency and Workload Reduction (f=6), Language Learning Enhancement (f=5), and Learner Engagement and Autonomy (f=4).

Among these sub-codes, Material Development and Resource Accessibility was the most frequently referenced (f=8). Participants emphasized that AI facilitates rapid access to educational resources and supports the preparation of instructional materials. Several instructors stated that AI enables teachers to generate materials tailored to students' individual needs within a short time. One participant noted that AI "provides rapid access to resources," while another stated that "materials tailored to students' individual needs can be generated quickly." Participants also underlined that AI can help teachers create instructional materials that would otherwise be difficult to prepare independently. These findings suggest that participants viewed AI as a practical tool for improving resource accessibility and supporting instructional design.

The second sub-code, Time Efficiency and Workload Reduction (f=6), indicates that participants perceived AI as a tool that can reduce routine

teaching workload and increase instructional efficiency. Instructors frequently associated AI with saving time in lesson preparation, planning, and content development. For instance, one participant stated that "artificial intelligence can save a considerable amount of time," while another noted that "it reduces teachers' workload and facilitates lesson preparation." Participants also emphasized that AI improves efficiency in instructional planning. These findings show that AI was perceived as a useful support mechanism for optimizing teachers' time and easing professional workload.

The third sub-code, Language Learning Enhancement (f=5), reflects participants' views on AI's potential contribution to students' language development. Participants stated that AI-supported activities can make lessons more engaging, modern, and interactive. One instructor noted that "artificial intelligence can make lessons more engaging and modern." Others emphasized that AI tools can support the development of speaking and reading skills and provide rapid feedback on writing and speaking tasks. These statements indicate that participants regarded AI as a tool that can enrich language learning experiences and support skill development across different areas of language education.

Finally, Learner Engagement and Autonomy (f=4) highlights participants' perceptions of AI as a means of encouraging active participation and independent learning. Participants suggested that AI can increase student engagement by offering interactive and personalized learning opportunities. One participant stated that "artificial intelligence can increase student engagement," while another emphasized that "students can manage their own learning process more independently." Participants also noted that AI creates opportunities for autonomous learning. These findings suggest that AI was perceived not only as a teaching aid but also as a tool that may help students take greater responsibility for their own learning.

Overall, the findings related to this theme indicate that participants generally perceived AI as a supportive educational technology. They associated AI with easier access to resources, more efficient material development, reduced workload, enriched language learning, and increased learner autonomy. However, these benefits appear most meaningful when AI is used under teacher guidance and integrated into teaching and learning processes in a purposeful and pedagogically appropriate manner.

Theme 2: Challenges and Risks Associated with Artificial Intelligence

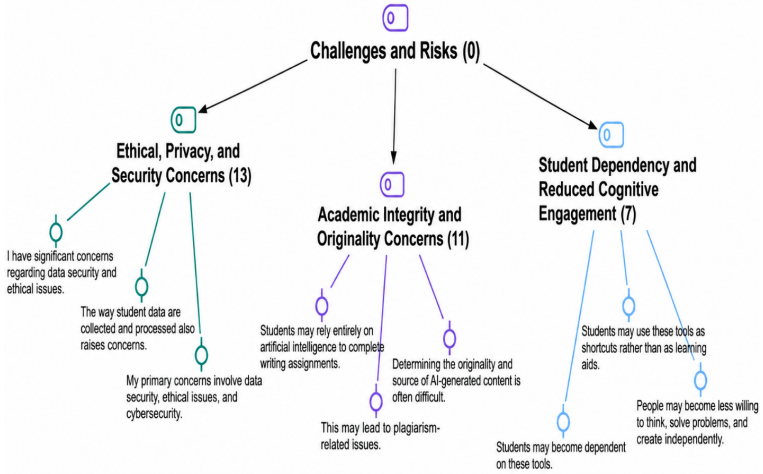


Figure 3. Theme 2: Challenges and Risks Associated with Artificial Intelligence (by MAXQDA)

Figure 3 presents the theme of Challenges and Risks Associated with Artificial Intelligence, which reflects participants' concerns about the potential difficulties that AI may create in educational settings. These concerns relate mainly to teaching, learning, assessment, and students' academic responsibility. Three sub-codes were identified under this theme: Ethical, Privacy, and Security Concerns (f=13), Academic Integrity and Originality Concerns (f=11), and Student Dependency and Reduced Cognitive Engagement (f =7).

Ethical, Privacy, and Security Concerns was the most frequently mentioned sub-code (f =13). Participants expressed serious concerns about the collection, storage, and processing of personal and student-related data in AI-supported educational environments. Several instructors referred specifically to data privacy, cybersecurity, and the ethical use of AI. One participant stated, "I have significant concerns regarding data security and ethical issues," while another noted that "the way student data are collected and processed also raises concerns." Similarly, another instructor emphasized that "my primary concerns involve data security, ethical issues, and cybersecurity." These statements suggest that participants viewed ethical governance, data protection, and privacy safeguards as essential requirements for the responsible use of AI in education.

The second sub-code, Academic Integrity and Originality Concerns (f=11), represents participants' concerns about the possible effects of AI on academic honesty and authentic student work. Participants frequently stated that AI-generated content may increase the risk of plagiarism and make it more difficult to assess students' original contributions. One participant noted that "this may lead to plagiarism-related issues," while another stated that "students may rely entirely on artificial intelligence to complete writing assignments." Participants also pointed to the difficulty of determining the source and originality of AI-generated content. As one instructor explained, "determining the originality and source of AI-generated content is often difficult." These findings indicate that participants considered academic integrity to be one of the major challenges associated with the increasing use of AI in educational contexts.

The third sub-code, Student Dependency and Reduced Cognitive Engagement (f = 7), focuses on participants' concerns about students' possible overreliance on AI tools. Participants argued that excessive dependence on AI may weaken students' critical thinking, problem-solving skills, creativity, and independent learning habits. One participant stated that "people may become less willing to think, solve problems, and create independently." Another participant noted that "students may become dependent on these tools," while a third observed that "students may use these tools as shortcuts rather than as learning aids." These statements show that participants were concerned that AI, when used without sufficient pedagogical guidance, may reduce students' active involvement in the learning process.

Overall, the findings under this theme show that participants identified several important risks related to the integration of AI into education. Ethical issues, data privacy, cybersecurity, academic integrity, originality, and student dependency were among the most prominent concerns. Although participants recognized the potential benefits of AI, they emphasized that its use should be guided by clear ethical principles, institutional safeguards, and teacher supervision. In this regard, AI should be integrated into educational settings in a way that supports meaningful learning while protecting students' academic responsibility, cognitive engagement, and originality.

Theme 3: Teacher Adaptation and Changing Professional Roles

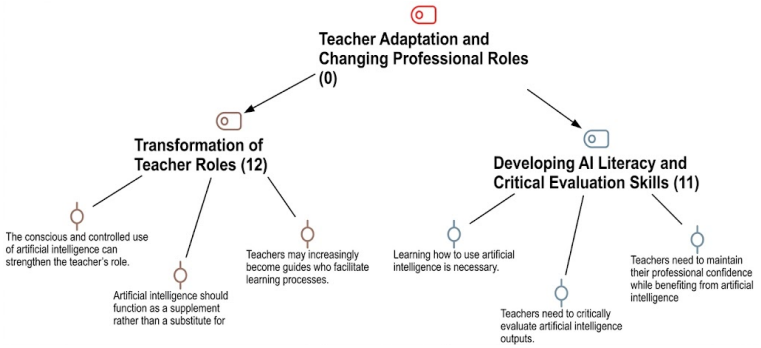


Figure 4. Theme 3: Teacher Adaptation and Changing Professional Roles (by MAXQDA)

Figure 4 presents the theme of Teacher Adaptation and Changing Professional Roles, which reflects how participants interpreted the professional adjustments required for the effective integration of AI into English language teaching. This theme consists of two sub-codes: Transformation of Teacher Roles (f = 12) and Developing AI Literacy and Critical Evaluation Skills (f = 11).

The first sub-code, Transformation of Teacher Roles, received the highest number of references (f = 12). Instructors emphasized that AI should not be understood as a technology that replaces teachers. Instead, they described it as a complementary tool that can enrich and support pedagogical practice. One participant stated that "artificial intelligence should function as a supplement rather than a substitute for teachers." Another participant similarly noted that "the conscious and controlled use of artificial intelligence can strengthen the teacher's role." Participants also drew attention to the changing nature of teachers' responsibilities in AI-mediated learning environments. As one instructor explained, "teachers may increasingly become guides who facilitate learning processes." These statements suggest that participants viewed AI as a factor that redefines, rather than weakens, the teacher's professional role. In this respect, teachers' traditional responsibility for transmitting knowledge is increasingly accompanied by new roles, such as guiding learners, organizing learning processes, and critically evaluating AI-generated content.

The second sub-code, Developing AI Literacy and Critical Evaluation Skills (f = 11), refers to the competencies teachers need in order to use AI responsibly and effectively. Participants stated that teachers should understand how AI tools work and how these tools can be used in pedagogically

meaningful ways. One participant clearly expressed this need by stating that "learning how to use artificial intelligence is necessary." However, participants did not define AI literacy only in terms of technical knowledge. They also emphasized the importance of critical judgment. As one instructor noted, "teachers need to critically evaluate artificial intelligence outputs." Another participant underlined the need for teachers to preserve their professional confidence while using AI, stating that "teachers need to maintain their professional confidence while benefiting from artificial intelligence." These views indicate that AI integration requires not only familiarity with digital tools, but also the ability to evaluate the accuracy, relevance, reliability, and instructional value of AI-generated responses.

Taken together, the findings related to this theme indicate that participants understood AI integration as a process that requires professional renewal and adaptation. Instructors did not position AI as a direct threat to the teaching profession. Rather, they viewed it as a development that reshapes teachers' responsibilities and increases the need for new professional competencies. The prominence of both sub-codes shows that the meaningful use of AI in education depends on teachers' ability to adapt to evolving roles, strengthen their AI literacy, and critically evaluate AI-generated content while maintaining their pedagogical expertise and professional identity.

Theme 4: Professional Development and Institutional Support Needs

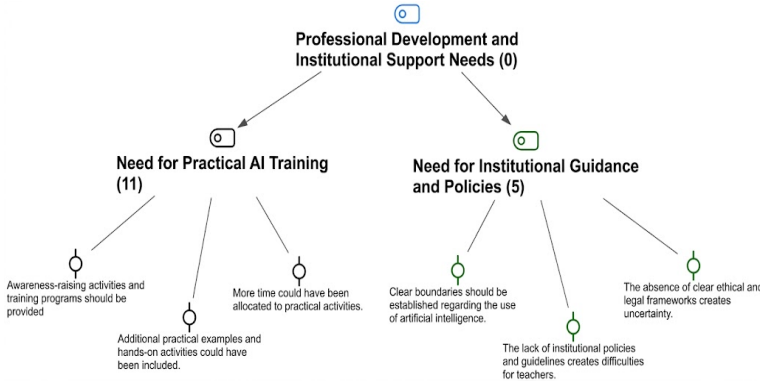


Figure 5. Theme 4: Professional Development and Institutional Support Needs (by MAXQDA)

As shown in Figure 5, the theme of Professional Development and Institutional Support Needs reflects participants' expectations regarding the conditions required for the effective, responsible, and sustainable use of AI in educational settings. Two sub-codes were identified under this theme: Need for Practical AI Training (f = 11) and Need for Institutional Guidance and Policies (f = 5).

The first sub-code, Need for Practical AI Training, was the most frequently emphasized category (f = 11). Participants stated that general awareness of AI is not sufficient for teachers to use these tools confidently in classroom contexts. Instead, they stressed the need for practice-based training that demonstrates how AI can be integrated into teaching activities in concrete and purposeful ways. One participant stated that "awareness-raising activities and training programs should be provided." Another participant noted that "additional practical examples and hands-on activities could have been included," indicating a clear expectation for more applied learning opportunities. Similarly, the statement that "more time could have been allocated to practical activities" shows that participants valued hands-on engagement with AI tools. These views suggest that instructors considered practical training a key condition for developing confidence, competence, and pedagogical control in AI-supported teaching.

The second sub-code, Need for Institutional Guidance and Policies (f = 5), concerns participants' expectations for clear institutional direction and regulatory support. Participants stated that the use of AI in education should be guided by explicit principles, boundaries, and ethical standards. One participant emphasized that "clear boundaries should be established regarding the use of artificial intelligence." This view points to the need for shared rules that clarify acceptable and responsible forms of AI use. Participants also expressed concern about the lack of formal guidance. One instructor stated that "the lack of institutional policies and guidelines creates difficulties for teachers," while another noted that "the absence of clear ethical and legal frameworks creates uncertainty." These statements indicate that participants associated institutional guidance with clarity, consistency, and reduced ambiguity in AI-related educational practices.

Viewed from this perspective, the theme shows that participants regarded professional development and institutional support as essential for meaningful AI integration. They did not view access to AI tools alone as sufficient. Rather, they emphasized the need for applied training, clear institutional policies, ethical guidelines, and organizational support. Therefore, the responsible use of AI in education appears to depend not only on

teachers' professional growth, but also on institutions' capacity to provide a clear, coherent, and supportive framework for implementation.

Relationships Among Themes and Codes

To examine the relationships among the themes and codes in greater detail, a code co-occurrence analysis was conducted using the MAXQDA Code Relations Browser, as presented in Figure 6.

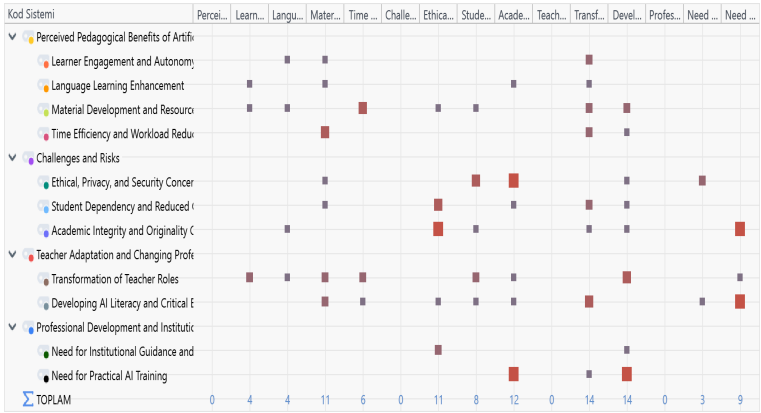


Figure 6. Co-occurrence analysis of Themes and Codes (by MAXQDA)

The Code Relations Browser shows how often different codes were assigned to the same segments of participant data. In this matrix, both the rows and columns represent the codes used in the analysis. Each square at the intersection of two codes indicates a co-occurrence between them. Larger and darker squares represent stronger relationships. Therefore, Figure 6 provides insight into how participants connected different dimensions of artificial intelligence use in English language teaching.

The analysis shows that participants did not discuss the opportunities and challenges of AI as separate issues. Instead, their responses often linked pedagogical benefits, ethical concerns, teacher adaptation, and professional support needs. This suggests that participants understood AI integration as a multidimensional process that affects teaching, learning, assessment, and institutional practice at the same time.

One of the strongest relationships in the figure was observed between Need for Practical AI Training and Developing AI Literacy and Critical Evaluation Skills. This co-occurrence indicates that participants associated AI literacy with practice-based professional learning. They did not view AI literacy as limited to theoretical knowledge. Rather, they

emphasized the need for hands-on training that helps teachers understand how AI tools work, evaluate AI-generated outputs critically, and use these tools in pedagogically appropriate ways. This finding suggests that practical training was regarded as an essential condition for responsible and meaningful AI integration.

Another strong relationship appeared between Ethical, Privacy, and Security Concerns and Academic Integrity and Originality Concerns. These codes frequently occurred together in participants' responses. This indicates that ethical issues were closely connected with concerns about plagiarism, authorship, originality, and the authenticity of student work. Participants seemed to view privacy, data security, and academic integrity as interrelated aspects of responsible AI use. This pattern also shows that AI-related risks were understood not only as technical or legal issues, but also as matters that directly influence assessment practices and educational trust.

A further strong relationship was identified between Academic Integrity and Originality Concerns and Need for Practical AI Training. This finding suggests that participants linked problems of academic integrity to the need for more concrete teacher training. Instructors appeared to believe that teachers need practical guidance on how to design assignments, evaluate student work, and prevent inappropriate AI use. In this respect, practical AI training was seen not only as a way to improve technical competence, but also as a means of supporting ethical and original student work.

The figure also shows a meaningful relationship between Transformation of Teacher Roles and Developing AI Literacy and Critical Evaluation Skills. This co-occurrence suggests that participants associated the changing role of teachers with the need to develop new professional competencies. As AI becomes more visible in language education, teachers are expected to act not only as knowledge providers, but also as guides, evaluators, and decision-makers. Therefore, the ability to assess the accuracy, relevance, and pedagogical value of AI-generated content becomes increasingly important. This finding indicates that teacher adaptation and AI literacy development were viewed as complementary processes.

Within the pedagogical dimension, a notable relationship was found between Material Development and Resource Accessibility and Time Efficiency and Workload Reduction. Participants often discussed these two issues together. This suggests that the perceived efficiency of AI was largely related to its ability to support material preparation, content generation, and access to instructional resources. Instructors appeared to value AI because it could reduce the time required for routine teaching tasks and support day-to-day instructional planning.

The analysis also shows a relationship between Ethical, Privacy, and Security Concerns and Student Dependency and Reduced Cognitive Engagement. This indicates that participants' concerns about AI were not limited to data protection or privacy. They also considered the possible effects of AI on students' learning responsibility, independent thinking, and cognitive engagement. From this perspective, uncontrolled AI use was perceived as a risk that may weaken students' active participation in the learning process.

Some weaker but still meaningful relationships were also visible among the pedagogical codes. For example, Language Learning Enhancement co-occurred with Learner Engagement and Autonomy. This suggests that participants sometimes connected AI-supported language development with increased student engagement and independent learning. However, this relationship was not among the strongest patterns in the figure. Therefore, it should be interpreted with caution.

The code Need for Institutional Guidance and Policies showed a more limited pattern of co-occurrence. It was mainly associated with Ethical, Privacy, and Security Concerns and, to a lesser extent, with Developing AI Literacy and Critical Evaluation Skills. This suggests that participants linked institutional guidance primarily to ethical clarity, data protection, and responsible teacher competence. The figure does not show a strong direct relationship between Need for Institutional Guidance and Policies and Need for Practical AI Training. Therefore, institutional support appears to be connected more directly with ethical regulation and AI literacy than with practical training alone.

Overall, the Code Relations Browser analysis indicates that participants conceptualized AI integration as an interconnected educational process. The strongest relationships were observed among practical training, AI literacy, academic integrity, and ethical concerns. Pedagogical benefits were mainly connected with material development and time efficiency. Teacher adaptation was closely related to the development of critical AI literacy. Institutional guidance was associated with ethical and responsible use. These findings show that instructors' perceptions of AI were not limited to a simple distinction between benefits and risks. Rather, participants viewed AI as a complex phenomenon that influences teaching practices, student responsibility, professional roles, ethical decision-making, and institutional structures simultaneously

Discussion

This qualitative case study examined English instructors' perceptions of ChatGPT and AI applications in foreign language teaching after a four-hour in-service training program. More specifically, it explored how instructors interpreted the contributions of AI, the risks and concerns associated with its use, the changing professional role of teachers, and the forms of professional and institutional support needed for responsible classroom use. The results point to a balanced but cautious understanding of AI. Participants did not frame AI as a simple solution to instructional problems. Rather, they viewed it as a powerful support mechanism whose value depends on teacher judgment, ethical awareness, and clear institutional guidance.

The first theme indicates that instructors perceived AI as a useful instructional support tool, particularly in relation to time efficiency, material development, language learning support, and learner autonomy. This finding supports the view that "AI technologies move beyond traditional approaches in foreign language teaching by offering more personalized, interactive, and task-based learning environments" (Dinar, 2025, p. 72). In the present study, instructors did not define AI merely as a tool for producing ready-made content. Rather, they emphasized its role in generating initial drafts, diversifying instructional options, and helping teachers respond more quickly to learners' needs. This result is consistent with studies suggesting that ChatGPT can support language learning through rapid feedback, personalized practice, writing assistance, and flexible access to resources (Barrot, 2023; Chen, 2023; Dizon, 2026; Karataş et al., 2024; Yüzlü, 2024). It also aligns with Aydin and Candemir Özkan's (2026) findings, which showed that learners perceived AI tools and environments as more supportive of writing and reading skills, while they adopted a more cautious stance regarding their effects on speaking, listening comprehension, pronunciation, and communication skills. In this respect, the present findings suggest that AI may contribute to different areas of language learning, although its perceived contribution may be more evident in text-based skills than in oral and interactional skills. The participants' emphasis on workload reduction is also in line with Büyüksaraç (2026) and Orhan and Cinar Yagci (2026), who note that ChatGPT can assist teachers in preparing lesson plans, activities, dialogues, and classroom materials. However, the present finding adds a context-specific nuance. For these four instructors, the value of AI was not limited to speed or convenience. Its instructional contribution became meaningful when AI-generated outputs were reviewed, adapted, and aligned with students' proficiency levels, course aims, and classroom realities. In this respect, AI

should be understood as a complementary resource that supports teacher decision-making rather than as an autonomous substitute for teacher expertise.

The second theme reveals that instructors were aware of the risks associated with AI use, especially academic integrity, originality, student dependency, reduced cognitive engagement, data security, and ethical uncertainty. These concerns are closely aligned with the literature, which warns that ChatGPT may produce unreliable, incomplete, decontextualized, or fabricated information and that its use in academic writing requires careful supervision (Barrot, 2023; Çınar Yağcı & Aydın Yıldız, 2023; Orhan & Cinar Yagci, 2026). The participants' concern about plagiarism and authorship also overlaps with Werdiningsih et al. (2024), who discuss the tension between AI-supported writing and authenticity in EFL contexts. Yet the instructors in the present study did not reject AI because of these risks. Their accounts suggest a more refined position: AI may support learning, but uncontrolled use may weaken students' effort, creativity, and independent thinking. This concern is important for foreign language education because language learning requires sustained cognitive engagement, repeated practice, error awareness, and gradual development. If students use AI mainly as a shortcut, the tool may reduce the very learning processes it is expected to support. At the same time, the risks should not be overstated. The participants' concerns point less to the inherent danger of AI itself and more to the need for supervised, transparent, and ethically guided use. In this sense, the findings support a responsible-use perspective rather than a prohibition-oriented approach.

The third theme shows that instructors interpreted AI as a factor that reshapes, rather than replaces, the teacher's professional role. Participants emphasized that teachers must become more active as designers, evaluators, facilitators, ethical guides, and critical decision-makers. This result is strongly compatible with Öztürk Demir and Çetin (2025), who underline the need for teacher guidance and conscious use of ChatGPT in foreign language teaching. It also extends the teacher-focused findings of Gao et al. (2024) and Orhan and Cinar Yagci (2026), which show that teachers make sense of AI through both its instructional benefits and its limitations. The emerging picture points to a shift from the teacher as the sole source of knowledge to the teacher as the person who evaluates the reliability, appropriateness, and educational value of AI-generated content. This shift does not reduce the importance of teachers. On the contrary, it makes their professional expertise more visible. Instructors must decide which outputs are accurate, which are suitable for learners, which may require revision, and which should not be used at all. Thus, AI use requires not only technical familiarity but also critical AI literacy. For

teacher education, this result is significant because future language teachers need to learn how to design AI-supported activities, guide students in responsible use, and maintain the human, relational, and ethical dimensions of teaching.

The fourth theme highlights instructors' need for practice-based professional development and clear institutional support. Participants valued the training program, but they expected more hands-on examples, classroom-based applications, and concrete guidance. This result extends previous work by showing that teachers' willingness to use AI is closely connected to practical competence and institutional clarity. Studies such as Yang and Li (2024) and Kır and Tekin (2025) emphasize that AI tools in language learning should be evaluated according to their purposes, uses, and contributions to learning. The present study adds that such evaluation cannot be left only to individual teachers. Instructors need institutional policies that clarify ethical boundaries, assessment principles, acceptable uses, data-related responsibilities, and student guidance procedures. Without such clarity, teachers may remain uncertain about how far they can go in using AI or how they should respond to student use. Practice-based training is therefore essential. It should include prompt design, material revision, feedback evaluation, assessment redesign, privacy awareness, and examples from real language classrooms. This need is especially important in higher education contexts where teachers are expected to respond quickly to technological change while also protecting academic integrity and learning quality.

The relationships among themes and codes suggest that instructors' perceptions of AI were interconnected rather than fragmented. Benefits such as time efficiency and material development were closely linked, indicating that teachers valued AI because it helped them produce and access resources more easily. Likewise, concerns about academic integrity were connected with fears of student dependency and reduced cognitive engagement. This overlap suggests that instructors did not see plagiarism only as a rule violation; they also viewed it as a threat to meaningful learning. Another important connection appeared between AI literacy and changing teacher roles. Teachers' ability to evaluate AI-generated content was central to how they understood their evolving responsibilities. Ethical and privacy concerns were also related to the need for institutional guidance, showing that responsible AI use cannot be reduced to individual awareness alone. Taken together, these relationships indicate that AI in foreign language teaching should be examined as a multidimensional issue involving instructional practice, teacher agency, learner responsibility, ethics, and institutional decision-making.

Overall, the findings suggest that AI in foreign language teaching should be understood neither as a purely beneficial innovation nor as a direct threat to education. Instead, it should be approached as a complex educational phenomenon that requires critical AI literacy, teacher agency, ethical awareness, and institutional support. ChatGPT may help teachers save time, design richer materials, support language practice, and provide feedback opportunities. However, these benefits become meaningful only when teachers review, adapt, and supervise AI use. Similarly, the risks of dependency, academic dishonesty, unreliable outputs, and data-related concerns can be managed more effectively when institutions provide clear policies and when teachers are prepared through practice-oriented training. Given the small number of participants and the qualitative case study design, these results should not be generalized to all English instructors. Nevertheless, the study offers a useful, context-sensitive account of how instructors in a Turkish higher education setting make sense of AI after professional training. The central implication is clear: responsible AI use in foreign language teaching depends less on the tool itself than on the professional, ethical, and institutional conditions under which it is used.

Conclusion

This study examined English instructors' perceptions of ChatGPT and AI applications in foreign language teaching after a four-hour in-service training program. The findings indicate that instructors approached AI with a balanced and reflective stance. They recognized its potential to support lesson preparation, material development, feedback practices, language skill development, and learner autonomy. At the same time, they did not view AI as an independent solution to instructional challenges. Rather, they considered it a supportive tool whose educational value depends on how it is selected, evaluated, adapted, and used within a meaningful teaching context.

Overall, the study suggests that the role of AI in foreign language teaching should be understood through the relationship between technological affordances and human judgment. ChatGPT and similar tools may reduce teachers' workload, broaden access to instructional resources, and create additional opportunities for practice. However, these benefits become educationally valuable only when they are guided by teacher expertise, ethical awareness, learner responsibility, and clear institutional direction. This interpretation is consistent with the view that teachers are no longer positioned merely as transmitters of knowledge; rather, they are actors who use technological tools within an ethical framework, organize learning processes, provide emotional support, and manage digital pedagogy

(Topaloğlu & Erol, 2025, p. 279). In this sense, the findings underline that responsible AI use is not merely a matter of adopting new tools. It requires critical AI literacy, careful instructional decision-making, and shared principles that protect the quality, integrity, and human dimension of language education.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the study was conducted with four English instructors working at the School of Foreign Languages of a single state university in Türkiye. The small number of participants allowed for an in-depth exploration of instructors' perceptions, but it limits the transferability of the findings to broader groups of language teachers or different educational contexts. Therefore, the results should be read as a context-sensitive account of instructors' experiences rather than as representative of all English language instructors.

A second limitation concerns the research design and data source. As a qualitative case study, the research relied on semi-structured interview data and therefore reflected participants' self-reported perceptions after the training program. The study did not include classroom observations, student data, teaching materials produced by participants, or long-term follow-up evidence. In addition, the training program lasted four hours, which means that participants' views were shaped by an initial professional development experience rather than by extended classroom use of AI tools. These limitations do not reduce the value of the study, but they suggest that the findings should be interpreted within the boundaries of a short-term, interview-based qualitative inquiry.

Recommendations

The findings point to several recommendations for foreign language teaching contexts. Institutions should offer practice-based AI training programs that move beyond general awareness and focus on real classroom applications. In this regard, it has been recommended that teachers' professional development processes should first be restructured so that AI technologies can be used effectively, ethically, and sustainably in foreign language teaching (Çınar Yağcı & Aydın Yıldız, 2025, p. 219). Such training should include prompt writing, material development, feedback evaluation, lesson adaptation, ethical use, and examples of responsible student engagement with AI tools. Teachers also need clear institutional guidelines that define acceptable uses of AI, address academic integrity, clarify privacy and data-related responsibilities, and support consistent

decision-making across courses. In addition, students should be guided to use AI as a learning aid rather than as a substitute for effort, reflection, and independent language production.

Future research should examine AI use in foreign language teaching with larger and more diverse participant groups across different institutions, proficiency levels, and educational settings. Longitudinal studies would be useful for understanding how teachers' perceptions and practices develop after sustained experience with AI-supported teaching. Mixed-method studies may also provide a broader understanding by combining interview data with surveys, classroom observations, document analysis, and learner performance data. Further research should include student perspectives, observed classroom practices, and redesigned assessment tasks in order to examine how AI affects learning responsibility, authorship, feedback practices, and the development of language skills over time.

Declaration on the Use of Generative Artificial Intelligence

In this study, generative artificial intelligence tools were used solely to support translation and linguistic editing processes. The theoretical framework, scientific content, source selection, and final interpretations of the study were developed, reviewed, and revised by the authors. Full scientific and ethical responsibility for the study rests with the authors.

References

Al-Maawali, W. (2026). Exploring Omani EFL teachers' perspectives on AI-generated cultural content in ELT. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2026.2669321

Ataş, A. H., Cengiz, B. C., & Çelik, B. (2024). Evaluating ChatGPT in Generating Feedback on Content and Organization Components of EFL Compare and Contrast Essays. Sakarya University Journal of Education, 14(3), 515-538. https://doi.org/10.19126/suje.1475474

Aydin, S. & Candemir Özkan, C. (2026). How Do EFL Learners Perceive Using Artificial Intelligence? In R. Zheng (Ed.), *The Intersection of AI and Student Development in Learning* (pp. 89-118). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-2927-7.ch003

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bayraktar Balkır, N., & Zehir Topkaya, E. (2025). AI in Foreign Language Learning: Changing Usage Practices, Perceptions, and Training Needs among EFL Students. Bogazici University Journal of Education, 42(3), 73-90. https://doi.org/10.52597/buje.1687891

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

Büyüksaraç, Z. (2026). Le Petit Chaperon rouge en classe de FLE : Confrontation des pratiques classiques et des productions de ChatGPT. *HAYEF: Journal of Education*, 23, 0086, doi:10.5152/hayef.2026.25086.

Chen, T. J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association, 86*(4), 351–353. https://doi.org/10.1097/JCMA.0000000000000900

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.

Çınar Yağcı, Ş. (2026). L'Intelligence Artificielle Generative Dans L'enseignement/Apprentissage Des Langues Etrangeres : Tendances, Principaux Resultats Et Perspectives De Recherche. In N. Özçelik (Ed.), Researches Internationales en Didactique du Français Langue Etrangere-I (pp. 49-95). Eğitim Yayınevi.

Çınar Yağcı, Ş. & Aydın Yıldız, T. (2023). ChatGPT, yabancı dil öğrencisinin güvenilir yapay zekâ sohbet arkadaşı mıdır? *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 37*, 1315–1333. https://doi.org/10.29000/rumelide.1407539

Çınar Yağcı, Ş. & Aydın Yıldız, T. (2025). Yabancı dil eğitiminde yapay zekâ kullanımı üzerine Türkiye'deki lisansüstü tezlerin meta-sentez yöntemiyle eleştirel incelenmesi (2020–2024). In Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (pp. 190–229). Günce Yayınları.

Çınar Yağcı, Ş. & Genç, A. (2024). Üretken yapay zekâ araçları ile Fransızca dijital hikâye tasarlama: Yazma, görselleştirme ve çeviri süreçleri. In N. Özçelik (Ed.), *Fransız dili eğitimi alanında uluslararası araştırmalar I / International research in the field of French language teaching* (pp. 85–110). Eğitim Yayınevi.

Dinar, G. (2025). Yapay zekâ ve eylem odaklı yabancı dil öğretimi: Kuram, uygulama ve gelecek perspektifleri. Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (s. 54–79). Günce Yayınları. ISBN: 978-625-5613-40-0

Dizon, G. (2026). ChatGPT as a tool for self-directed foreign language learning. *Innovation in Language Learning and Teaching*, *20*(2), 334–350. https://doi.org/10.1080/17501229.2024.2413406

Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: A study in China. *Asia Pacific Journal of Education, 44*(1), 29–44. https://doi.org/10.1080/02188791.2024.2305173

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*, 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's

effect on foreign language learners. *Education and Information Technologies*, *29*(15), 19343-19366.

Kır, D. B., & Tekin, E. (2025). Key Features of AI Applications in Foreign Language Teaching. Bartın University Journal of Faculty of Education, 14(4), 1143-1162. https://doi.org/10.14686/buefad.1681050

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage.

Orhan, A., & Cinar Yagci, Ş. (2026). Using ChatGPT in EFL context: What do EFL teachers think about it? *International Journal of Technology in Education, 9*(2), 447–475. https://doi.org/10.46328/ijte.5323

Özçelik, N., & Çınar Yağcı, Ş. (2024). L'enseignement des langues étrangères soutenu par l'intelligence artificielle générative: L'effet de ChatGPT sur l'amélioration des compétences rédactionnelles en français. In N. Özçelik (Ed.), Fransız dili eğitimi alanında uluslararası araştırmalar I (pp. 25-60). Eğitim Yayınevi.

Özdemir Çiftçi, S. Ö., & Tilbe, A. (2025). TPACK modeli bağlamında Fransızca öğretiminde ChatGPT'nin kullanımı: Dört temel dil becerisi ve pedagojik roller üzerine bir inceleme. In Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (pp. 158–189). Günce Yayınları.

Öztürk Demir, S., & Çetin, M. (2025). Bir Üretken Yapay Zekâ Aracı Olarak ChatGPT'nin Yabancı Dil Öğretiminde Kullanılmasına İlişkin Görüşler. Dil Eğitimi ve Araştırmaları Dergisi, 11(1), 207-234. https://doi.org/10.31464/jlere.1501596

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage.

Sarı, Y. E. (2025). Üretken yapay zekâ ile beceri temelli Almanca eğitimi. In Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (pp. 230–251). Günce Yayınları.

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2026). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, *39*(4), 1148–1175. https://doi.org/10.1080/09588221.2025.2454541

Topaloğlu, Y., & Erol, Ü. (2025). Yapay zekâ ile dönüşen yabancı dil eğitimi: Kişiselleştirme, katılım, ekinsel yeterlilik ve etik sınırlar bağlamında eğitsel/öğretimsel bir yeniden yapılanma. In Y. Topaloğlu (Ed.), *Yapay zekâ ve yabancı dil eğitimi: Kuramsal temeller, uygulamalar ve yeni yaklaşımlar* (pp. 252–288). Günce Yayınları.

Üstünbaş, Ü. (2024). EFL Learners' Views About the Use of Artificial Intelligence in Giving Corrective Feedback on Writing: A Case Study. In Z. Çetin Köroğlu & A. Çakır (Eds.), *Fostering Foreign Language Teaching and Learning Environments With Contemporary Technologies* (pp. 115-133). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-0353-5.ch006

Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities, 11*(1), 1–15. https://doi.org/10.1080/23311983.2024.2392388

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages, 8*, 1–12. https://doi.org/10.3390/languages8030212

Yalçın, P., & Çınar Yağcı, Ş. (2024). Fransız Kültür Edincinin Kazandırılmasında Yapay Zekâ Rehber Olabilir Mi? Bing Chat AI Örneği. *Yükseköğretim Ve Bilim Dergisi,* 14(2), 308-327. https://doi.org/10.5961/higheredusci.1402198

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, *124*, 103351.

Yıldız, C. (2024). ChatGPT Integration in EFL Education: A Path to Enhanced Speaking Self-Efficacy. Novitas-ROYAL (Research on Youth and Language), 18(2), 167–182. https://doi.org/10.5281/zenodo.13861137

Yüzlü, M. Y. (2024). The Impact of ChatGPT on Vocabulary Learning in L3 German Class. Yükseköğretim Ve Bilim Dergisi, 14(2), 339-348. https://doi.org/10.5961/higheredusci.1417301

The Transformation of the Teacher's Role in the Age of Artificial Intelligence: A Didactic Decision Tree Proposal Based on a Prompt Engineering Taxonomy

Hakan Dolan [1]

Abstract
Artificial intelligence is reshaping teachers' professional roles and increasing the need for pedagogically grounded AI competencies. This conceptual study examines prompt engineering as an emerging didactic competence and proposes a decision-support framework for foreign language education. Drawing on current literature on AI literacy, teacher competencies, and prompt engineering, the study reinterprets the 41-technique taxonomy according to instructional needs in foreign language teaching. Particular attention is given to selecting techniques for reliable, source-grounded material development. A scenario-based decision tree is developed to distinguish among source-grounded/document-grounded prompting, Retrieval-Augmented Generation (RAG), Chain-of-Note (CoN), Chain-of-Knowledge (CoK), ReAct, and Chain-of-Verification (CoVe) according to source structure, information needs, and external tool use. The framework positions these techniques as complementary pedagogical decision-support mechanisms rather than automated substitutes for teacher expertise. It further emphasizes that model-assisted verification must be followed by mandatory teacher review to ensure linguistic accuracy, source fidelity, proficiency-level appropriateness, cultural validity, and pedagogical usability.
Keywords: *artificial intelligence, prompt engineering, foreign language education, teacher competencies, decision tree*

Introduction

Although many teachers are willing to utilize AI, their personal experiences and practical knowledge remain limited; hence, there is a need for AI-focused professional development programs with specific objectives (Kalimullina et al., 2020; D. Ng et al., 2023; Tenberga and Daniela, 2024; Velander et al., 2024). Because AI literacy builds partly on existing digital competencies while also requiring AI-specific knowledge and skills, targeted professional development is important for supporting teachers' meaningful integration of AI into educational practice (Tenberga & Daniela, 2024). One of the most critical skills necessary for the effective utilization of AI technologies is prompt writing. The ability to formulate clear, goal-oriented, and well-structured prompts is a critical skill for the

[1] Lecturer, National Defence University, School of Foreign Languages, Department of Western Languages, German Language Program, ORCID: 0000-0001-9644-2644, *hakan.dolan@msu.edu.tr*, Türkiye

effective use of generative AI, as prompt quality can directly influence the quality of generated outputs and creative problem-solving performance (Knoth et al., 2024; Urban et al., 2025).

Prompt engineering emerges as a strategic paradigm that guides the outputs of large language models through task-specific instructions without modifying internal model parameters. The prompt engineering taxonomy established by Sahoo et al. (2025) offers a comprehensive framework for this domain by classifying numerous techniques, including zero-shot, few-shot, Chain-of-Thought, Retrieval-Augmented Generation, ReAct, Chain-of-Verification, Chain-of-Note, Chain-of-Knowledge, Emotion Prompting, and Take a Step Back, based on their specific application fields. However, this taxonomy does not provide a pedagogical decision model specifically designed for foreign language education. In response, the present study reinterprets this taxonomy in accordance with the instructional requirements of foreign language education, thereby proposing pedagogical objective-based decision trees designed to assist foreign language educators in selecting pedagogically appropriate prompt engineering techniques for specific instructional goals. Accordingly, the study conceptualizes prompt engineering not merely as a technical AI skill but as a pedagogical decision-support mechanism designed to support material development, source-grounded content generation, formative feedback, grammar instruction, interactive practice, and assessment.

The Teacher's Role in the Process of Digital Transformation

In tandem with digital transformation, teachers have increasingly been positioned as transformative agents and key actors in digital leadership and school-level innovation (Antonopoulou et al., 2025; Reinius et al., 2022). Digital transformation shifts the role of the teacher from a mere transmitter of knowledge into a multidimensional professional who designs, executes, and evaluates digital learning processes. Recently, artificial intelligence technologies, which represent one of the most prominent advancements in digital transformation, have elevated teachers to the status of learning facilitators, designers, and data-driven decision-makers (Chen, 2025; Aspandi & Muttaqin, 2025). AI-enabled learning analytics and adaptive systems can provide educators with information about learners' progress, needs, and interaction patterns, which can inform personalized instructional support and course redesign while also supporting students' self-regulated learning (Ren & Wu, 2025). The following table illustrates the specific dimensions through which this transformation grants new roles to teachers within the context of artificial intelligence.

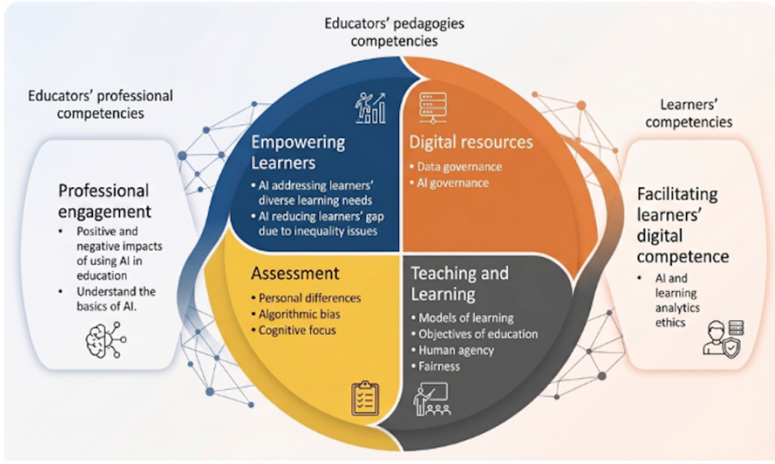
Table 1. Dimensions of the teacher's role in the AI era

Core Roles Transformed by AI		
Role Dimension	**Emphasis in the Context of AI**	**Relevant Studies**
Instructional Designer	AI-supported personalized learning pathways	Antonopoulou et al., 2025; Ren & Wu, 2025; Aspandi and Muttaqin, 2025
Data-driven Decision Maker	Intervention through the interpretation of learning analytics	Antonopoulou et al., 2025; Ren & Wu, 2025
Ethical Guide / AI Literacy Facilitator	Facilitating discussions on bias, transparency, and accountability	Markauskaite et al., 2023; Chiu et al., 2024; Sperling et al., 2024;
Digital Leader / Change Agent	Transforming school culture and practices	Reinius et al., 2022; Wohlfart & Wagner, 2023; Yue et. al., 2024; Tenberga & Daniela, 2024

Artificial Intelligence Literacy and Competency as Emerging Teacher Competencies

AI literacy generally encompasses the knowledge and skills required to understand, critically evaluate, and responsibly engage with AI, whereas AI competency extends this foundation to the confident, reflective, and contextually appropriate application of such knowledge and skills (Chiu et al., 2024; Sperling et al., 2024). Owing to the positive correlation identified between educator digital competence and attitudes toward AI (Galindo-Domínguez et al., 2024), various studies define pedagogical AI competencies for teachers as multidimensional constructs comprising content knowledge and AI integration across learning, teaching, assessment, and evaluation (Mikeladze et al., 2024; Zou et al., 2025). The European Commission's (2022) DigCompEdu Framework for Artificial Intelligence Competencies designed for educators serves as a foundational benchmark specifically for AI integration in teacher education and research within the domain of educational technologies.

Figure 1. Conceptual relationship between DigCompEdu domains and AI-related teacher competencies. Developed by the authors based on Redecker (2017), European Commission (2022), and Ng et al. (2023)



This illustration demonstrates that educators' digital competencies in the era of artificial intelligence cannot be reduced to the mere utilization of technological tools. Rather, these competencies must be addressed in tandem with multidimensional capacities such as pedagogical decision-making, ethical reasoning, data governance, learning analytics, human autonomy, and digital citizenship. The core structure of the DigCompEdu framework mandates that teachers be digitally competent across the domains of professional engagement, digital resources, teaching and learning, assessment, learner empowerment, and facilitating learners' digital competence (Redecker, 2017; Caena & Redecker, 2019). The European Commission's (2022) ethical guidelines complement this digital competence perspective by emphasizing educators' ability to critically evaluate AI systems, understand their ethical implications, and make informed decisions regarding the use of AI and data in teaching and learning.

The professional engagement domain, located on the left side of the illustration, signifies the educator's critical evaluation of both the positive and negative impacts of AI in education, their understanding of the fundamental mechanics of AI and learning analytics, and their active participation in continuous professional development. This dimension mandates that teachers conceptualize AI systems not merely as functional tools but as socio-technical systems embedded with pedagogical and ethical implicationsThis perspective requires educators to move beyond the functional

use of AI tools and to critically consider their pedagogical, societal, and ethical implications (European Commission, 2022; Long & Magerko, 2020). Similarly, UNESCO's AI competency framework for teachers states that the teacher-student dynamic in the AI era has transitioned into a teacher-AI-student interaction, thereby necessitating a comprehensive redefinition of educator roles (Miao & Cukurova, 2024). The central digital resources domain demonstrates that the selection, creation, and sharing of digital resources within AI-supported educational environments must now be conceptualized alongside data governance and AI governance. Within this context, educators are expected to possess a foundational awareness regarding personal data processing, data security, data retention periods, the role of third-party service providers, and automated decision-making processes (European Commission, 2022). Furthermore, the teaching and learning domain emphasizes the role of the teacher as a pedagogical designer in integrating AI systems into instructional processes. The inclusion of concepts such as 'models of learning,' 'objectives of education,' 'human agency,' and 'fairness' within the illustration reveals that AI systems are not neutral instruments. Instead, they incorporate specific models concerning the definition of learning, how it is assessed, and the pedagogical assumptions through which it is supported (European Commission, 2022). Consequently, when utilizing AI-backed learning systems, educators must conduct evaluations that center on human autonomy, pedagogical objectives, and learner well-being (Miao & Cukurova, 2024, UNESCO, 2021).

The assessment domain highlights critical issues in AI-based assessment and evaluation processes, including individual differences, algorithmic bias, and cognitive focus. AI systems often evaluate student progress based on predefined knowledge models, which creates the potential for underrepresenting more complex learning outcomes such as creativity, collaboration, social skills, or contextual performance (European Commission, 2022). Furthermore, algorithmic bias can reproduce or reinforce existing inequalities in educational technologies. Consequently, educators must critically interpret automated feedback, learning analytics, and predictive systems (Baker and Hawn, 2022). The empowering learners domain, while emphasizing the opportunities AI offers for personalized learning, inclusivity, and student engagement, also cautions that these opportunities may not yield equitable outcomes across all student populations. The European Commission notes that personalized learning systems are adaptive regarding content, learning pathways, and pedagogical approaches. Nevertheless, these systems can impact diverse student groups differently (European Commission, 2022). Therefore, the role of the educator is not to blindly adopt AI-supported personalization, but rather to

evaluate the system's pedagogical rationale, evidence-based value, and potential exclusionary effects (Ng et al., 2023; Salas-Pilco et al., 2022).

Lastly, the facilitating learners' digital competence domain demonstrates that educators must guide students not merely to use AI tools, but rather to comprehend the ethical, societal, and data-driven implications of these technologies. This dimension conceptualizes AI literacy as a broader concept than technical skill, thereby requiring individuals to critically evaluate AI systems, interact effectively with AI, and utilize these systems responsibly (Long and Magerko, 2020). This approach directly correlates with the principles of transparency, fairness, accountability, privacy, and responsibility frequently emphasized in the AI ethics literature (Jobin et al., 2019; UNESCO, 2021).

Therefore, the illustration conceptualizes teacher competence in the AI era at three foundational levels. The first level involves the teacher contextually structuring AI-supported learning processes as a pedagogical designer. The second level entails critically and ethically interpreting learning analytics as a data-driven decision-maker. The third level focuses on supporting students in engaging in conscious, safe, and responsible interactions with AI systems as an ethical guide and digital citizenship facilitator. In this regard, the framework transforms the role of the teacher from a passive practitioner who merely utilizes technology into an active agent of change who safeguards educational values, human autonomy, and societal responsibility (European Commission, 2022; Miao and Cukurova 2024; Regulation [EU] 2024/1689, 2024).

Prompt Engineering as a Didactic Competence

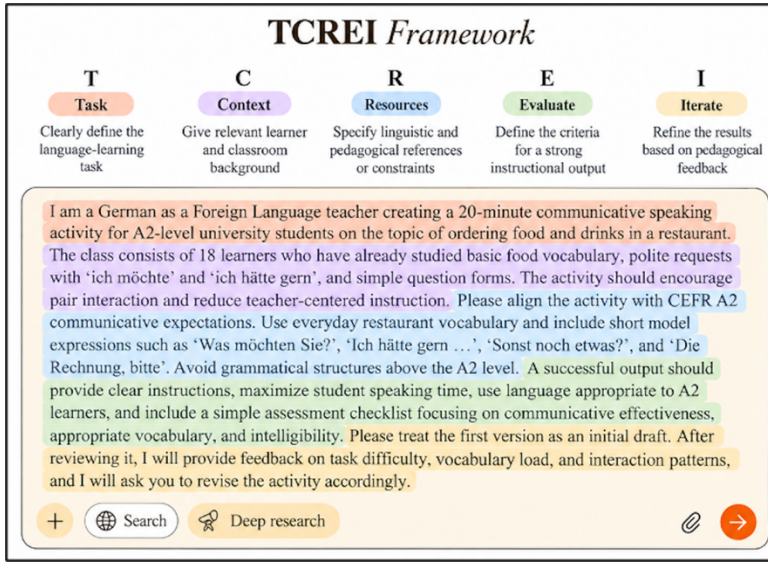
Prompt engineering is defined as a rapidly emerging didactic competence necessary for establishing effective communication with artificial intelligence tools and securing pedagogically meaningful outputs from these instruments (Lo, 2023; Walter, 2024; Elsayary et al. 2025; Leung, 2024; Qian, 2025; Dourvas et al., 2026). Recent systematic reviews and empirical studies demonstrate that prompt engineering is not merely a technical skill, but rather a holistic competence that supports the core pedagogical functions of educators, such as critical thinking, content design, assessment, and the creation of personalized learning environments (Cain, 2023; Leung, 2024). This competence encompasses the dimensions of knowledge regarding prompt types and techniques, skills involving effective prompt writing and evaluation, and attitudes focusing on a critical approach and ethical awareness (Elsayary, 2025; Abaidoo, 2026; Akhmadullina et al., 2026).

For educators, prompt engineering is utilized in core didactic processes such as lesson planning, material development, generating assessment questions, and designing personalized learning scenarios (Leung, 2024; Rey, 2025; Saefurrohman et al., 2026). Through effective prompts, teachers can foster critical thinking among students, stimulate higher-order cognitive processes, and promote collaborative learning (Lo, 2023; Walter, 2024; Leung, 2024). Evidence from language teacher education and vocational education indicates that structured training in generative AI and prompt engineering can enhance teachers' competencies in using AI for lesson planning and developing effective prompts (Huang et al., 2025; Schmitt & Brutzer, 2025). Recent reviews further suggest that structured prompting strategies may support higher-order cognitive processes such as analysis, reflective reasoning, and metacognitive engagement (Moguš, 2025). Furthermore, educators who receive prompt engineering training within continuous professional development programs experience not only an increase in lesson planning efficiency but also an elevation in material quality (Saefurrohman et al., 2026).

Frameworks for Prompt Engineering

Various frameworks have been developed to address prompt engineering as a didactic competence, including models such as the CLEAR Framework, comprising Concise, Logical, Explicit, Adaptive, and Reflective dimensions, and the TCREI model, consisting of Task, Context, Resources, Evaluate, and Iterate (Oreški et al., 2025). These frameworks integrate the dimensions of cognitive knowledge or skills, technical tool utilization, and ethical practice (Akhmadullina, 2026). The following illustration presents an adapted pedagogical example of the TCREI framework, demonstrating how it can be used to enhance prompt quality in human–AI interaction and to support more accurate outputs generated by large language models (LLMs).

Figure 2. TCREI Framework Sample for Foreign Language Education (adapted by the author using generative AI tools)



The color coding embedded in the illustration demonstrates through a pedagogical approach how a complex natural language command can be decomposed into its functional components. In the present example, the prompt is organized around the design of a communicative speaking activity in German as a Foreign Language, thereby illustrating how pedagogical aims, learner context, linguistic resources, evaluation criteria, and iterative refinement can be systematically integrated into prompt construction. Accordingly, prompting frameworks may contribute not only to improved output quality, but also to more pedagogically aligned lesson design, feedback generation, and instructional decision-making (Serra & Oliveira, 2025; Lee & Palmer, 2025; Alyahyan, 2025).

Prompt Taxonomy

Prompt engineering is identified as an indispensable and transformative technique utilized to enhance and extend the capabilities of large language models (LLMs). Fundamentally, it represents the process of eliciting desired behaviors from a model solely through task-specific instructions termed 'prompts,' without modifying or updating the core parameters of the model (Sahoo et al.,

2025). Whereas traditional artificial intelligence approaches require model retraining or extensive fine-tuning to introduce a new task, prompt engineering eliminates this requirement through strategic instruction design, thereby enabling models to exhibit high adaptability across diverse tasks and domains. Prompt engineering techniques represent methods that are more complex and purposefully diversified compared to standard prompt frameworks. Sahoo et al. (2025). established a comprehensive taxonomy by classifying these prompt engineering techniques according to their specific objectives and demonstrating empirical case studies for each method. Organized based on application domains, the classification of prompt engineering techniques utilized in large language models (LLMs) provides a detailed framework for customizing prompts across various pedagogical and instructional contexts.

Figure 3. Taxonomy of prompt engineering techniques in LLMs



Note. adapted from Sahoo et al. (2025)

The taxonomy delineates a total of 41 distinct prompt engineering techniques alongside empirical case studies illustrating their respective applications. These techniques are categorized based on their application fields and functional objectives, spanning a comprehensive spectrum that ranges from foundational methods such as zero-shot prompting to highly sophisticated approaches like Chain of Code.

Prompt engineering originates with foundational techniques such as Zero-Shot Prompting, which facilitates the adaptation of models to novel tasks without requiring training data, and Few-Shot Prompting, which enhances performance by presenting the model with several task exemplars. Over time, these have evolved into significantly more potent variants including Chain-of-Thought (CoT) Prompting, which enables the model to solve complex problems step-by-step in a logical sequence, and Automatic Chain-of-Thought (Auto-CoT) Prompting, which minimizes human intervention. Additional advanced iterations encompass Self-Consistency, which evaluates multiple valid reasoning pathways, and Logical Chain-of-Thought (LogiCoT) Prompting, which isolates erroneous steps through symbolic logic. Furthermore, to model the non-linear architecture of human cognition, Tree-of-Thoughts (ToT) Prompting utilizing branching hierarchies and the highly versatile Graph-of-Thoughts (GoT) Prompting have been developed. The reasoning capacity of models has been extensively expanded through specialized methods such as Chain-of-Table Prompting for tabular operations, Chain-of-Symbol (CoS) Prompting to overcome word-based ambiguities in text, and Self-Refine Prompting, wherein the model autonomously critiques and rectifies its initial output.

To mitigate the issue of models generating unrealistic hallucinations and to leverage external knowledge repositories, a distinct specialized category has been developed. Within this domain, pioneering methods include Retrieval-Augmented Generation (RAG), which retrieves real-time documents from external databases to provide context for the model, and ReAct Prompting, which executes reasoning and external actions simultaneously, such as utilizing the Wikipedia API. Reliability is further advanced through Chain-of-Verification (CoVe) Prompting, wherein the model validates its own generated responses by planning verification questions, Chain-of-Note (CoN) Prompting, which enables the model to output an unknown status when retrieved external information is irrelevant, and Chain-of-Knowledge (CoK) Prompting, which dynamically aggregates information from multiple sources rather than relying on a static database. Additionally, fine-tuning and optimization workflows have been completely re-engineered through Active Prompting, which utilizes active

learning strategies to focus on ambiguous instances that the model requires most, Automatic Prompt Engineer (APE), which dynamically generates optimal instructions via reinforcement learning, and Automatic Reasoning and Tool-use (ART), which integrates external instruments such as calculators into the process as automated programmatic routines.

Innovative prompt engineering designs focusing on more complex mathematical problems, code generation, emotion regulation, and metacognition are also available. To enable the model to learn not only from correct reasoning but also from erroneous logical steps, Contrastive Chain-of-Thought (CCoT) Prompting has been developed, whereas Emotion Prompting, which enhances output quality by embedding psychological expressions such as emphasizing the importance of the task into prompts, introduces a novel dimension to consistency. In coding tasks, logical precision is maximized through techniques such as Scratchpad Prompting, wherein the model utilizes a buffer memory for temporary computations, Program of Thoughts (PoT) Prompting, which translates complex numerical calculations into directly executable Python code rather than natural language, and Chain-of-Code (CoC) Prompting, which transforms even semantic subtasks into pseudocode. Lastly, methods such as Rephrase and Respond (RaR) Prompting, in which the model reformulates the question to align with its own semantic framework prior to generating an answer, Take a Step Back Prompting, which facilitates a focus on high-level abstract principles rather than specific instances, and Optimization by Prompting (OPRO), which effectively converts models into optimization engines through natural language instructions, collectively deepen the capacity of artificial intelligence to comprehend user intent and cultivate advanced insights.

Adaptation of Prompt Engineering Techniques to Foreign Language Education

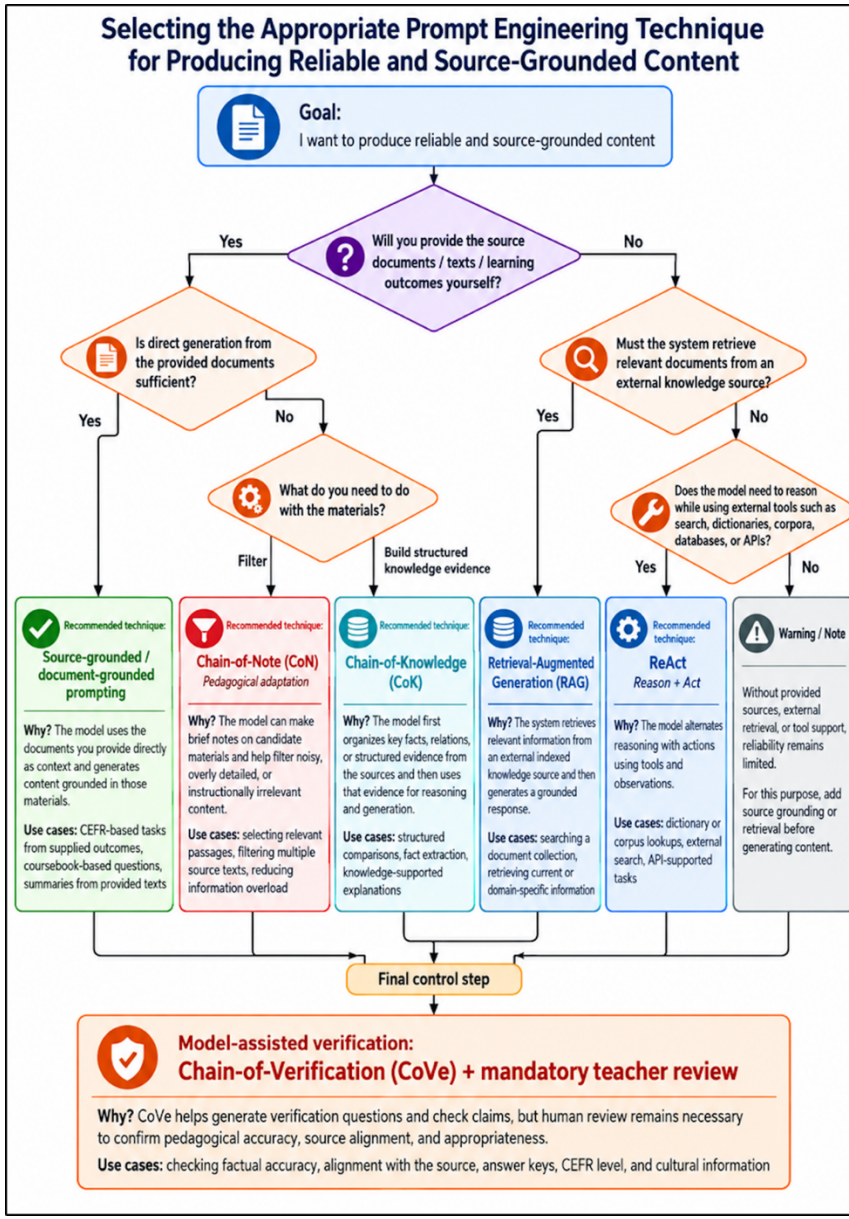
Current prompt engineering techniques that enhance the performance of large language models (LLMs) are predominantly structured around logical inference, algorithmic computation, and general knowledge-based question-answering tasks, having been optimized on benchmarks specific to these domains. Upon examining this taxonomy established in the literature, it becomes evident that the aforementioned techniques are not directly ready for deployment in a specialized field such as foreign language education and learning, which possesses its own unique pedagogical and linguistic dynamics.

The primary function of prompt engineering resides in the strategic design of task-specific instructions. Consequently, for existing techniques such as Chain-of-Thought, Retrieval-Augmented Generation, and Emotion Prompting to be successful in foreign language instruction, a structural adaptation based on targeted foreign language skills and specific learning objectives is required rather than a direct transfer. Furthermore, given the persistent challenges that models continue to harbor, including biases, factual inaccuracies, and interpretability gaps, it is critical that these techniques undergo context-specific modifications, various optimizations, and compliance screenings when being integrated into a highly sensitive domain like education.

Developing Reliable and Source-Grounded Lesson Content Through Scenario-Based Prompting

Within the field of foreign language education, a primary benefit of artificial intelligence resides in its capacity to offer alternatives tailored to individual learning needs by developing materials aligned with diverse learning styles and paces (Kessler, 2018). This section of the study elucidates how decisions can be made regarding the selection of prompt engineering techniques featured in the taxonomy through a specific scenario. In this scenario, a German language teacher intends to design a reading activity themed ‘Sustainable Transportation in Germany’ for B1-level students. The educator possesses CEFR B1 learning outcomes, a brief text from the course textbook, and two additional resources related to the topic. The objective of the teacher extends beyond merely generating a rapid text. The core aim is to construct an instructional material that is source-grounded, level-appropriate, pedagogically viable, and fully verified. To achieve this objective, the following decision tree has been designed and its progressive stages have been explained.

Figure 4. A Decision Tree for Selecting Appropriate Prompt Engineering Techniques for Source-Grounded Content Generation



Note. adapted by the author using generative AI tools

This decision tree re-functionalizes specific techniques from the prompt engineering taxonomy presented by Sahoo et al. (2025), most notably Retrieval-Augmented Generation, Chain-of-Note, Chain-of-Knowledge, ReAct, and Chain-of-Verification, within the context of foreign language education.

Stage 1: Source-Grounded Prompting / Document-Grounded Prompting

The initial question of the decision tree is formulated as follows: "Will the teacher provide the source documents, texts, learning outcomes, or other relevant materials?" In the present scenario, the answer is affirmative because the teacher directly provides the CEFR B1 learning outcomes, the textbook passage, and supplementary resources to the model. Since these materials are supplied directly rather than retrieved from an external indexed knowledge source, the appropriate approach at this stage is source-grounded or document-grounded prompting, rather than Retrieval-Augmented Generation (RAG). Source-grounded prompting constrains the generation process by explicitly requiring the model to rely on the materials supplied within the interaction. This distinction is particularly important in foreign language education, where instructional content may need to conform closely to curriculum objectives, proficiency-level descriptors, textbook materials, or culturally sensitive information. Accordingly, the teacher may use a prompt such as the following:

> You are an experienced German language teacher. Design a reading activity for B1-level students based strictly on the CEFR B1 learning outcomes and the textbook passage provided below. Rely exclusively on the supplied materials. Do not introduce factual information that cannot be supported by these sources. Provide a short reading passage, five comprehension questions, and an answer key.

In this context, source-grounded prompting serves three primary functions. First, it anchors the generated content to specific instructional materials. Second, it supports alignment between the output, curriculum objectives, and learner proficiency level. Third, it reduces the introduction of unsupported information beyond the materials supplied by the teacher. If the required information is not directly provided and must instead be retrieved from an external indexed source, the workflow shifts to Retrieval-Augmented Generation (RAG), as represented by the separate branch of the decision tree.

Stage 2: Pedagogical Adaptation of Chain-of-Note for Filtering Noisy or Irrelevant Materials

In the second stage, the teacher determines whether the available materials require filtering before they can be used for instructional content generation. In the present scenario, one resource contains excessive detail, whereas another includes information that is only partially relevant to the instructional objective. Under these circumstances, direct generation may result in unnecessary or pedagogically inappropriate content. The present framework therefore pedagogically adapts the logic of Chain-of-Note (CoN) to support the evaluation and filtering of candidate materials. Rather than treating every source fragment as equally useful, the model is instructed to produce concise notes identifying information that is relevant, usable, and appropriate for the target proficiency level. A prompt may be structured as follows:

> Analyze the three sources provided below. For each source, identify only the information that directly contributes to the topic of sustainable transportation for B1-level students. Exclude information that is irrelevant, unnecessarily detailed, or linguistically inappropriate for the B1 level. Present the selected information as concise notes that can subsequently be used to develop the instructional material.

Within this pedagogical adaptation, CoN does not directly generate the final instructional material. Instead, it functions as an intermediate filtering mechanism that helps reduce informational noise and identify source elements relevant to the instructional objective. This may be particularly useful when teachers work with extensive, heterogeneous, or partially irrelevant materials.

Stage 3: Using Chain-of-Knowledge to Build Structured Knowledge Evidence

The third stage concerns situations in which the teacher requires more than simple filtering and needs the model to organize relevant information into **structured knowledge evidence** before generating an instructional response. In the present scenario, selected information from the CEFR outcomes, textbook material, and supplementary sources may contain facts, concepts, and relations that first need to be explicitly organized. At this stage, Chain-of-Knowledge (CoK) is used not merely to "combine several sources," but to establish an explicit knowledge structure that can subsequently support reasoning and generation. The teacher may therefore use a prompt such as:

> Using the relevant information identified in the previous stage, extract and organize the key facts, concepts, and relationships concerning sustainable transportation. For each knowledge element, indicate the source from which it is derived. Identify any conflicting or unsupported claims. Then use only the resulting structured knowledge evidence to prepare the basis for a B1-level instructional text.

The central function of CoK in this framework is therefore to establish a transparent knowledge basis before final content generation. Within foreign language education, such structured evidence may support tasks involving factual explanation, cultural comparison, knowledge-supported reading materials, and instructional activities in which the relationship between source information and generated content must remain traceable.

Stage 4: Implementing ReAct When Actual External Tools Are Required

In the fourth stage, the decision concerns whether the model must interact with **actual external tools or information systems** while performing the task. These may include search engines, dictionaries, corpora, databases, or APIs. In the present scenario, the teacher seeks authentic examples of expressions such as *nachhaltige Mobilität*, *öffentliche Verkehrsmittel*, and *Fahrradnutzung*. If the AI system has genuine access to an external dictionary or corpus, a ReAct-based workflow may be appropriate. ReAct combines reasoning with actions performed through accessible external tools and incorporates the resulting observations into subsequent reasoning. Consequently, merely instructing a model to "think like a dictionary" or "use corpus logic" does not constitute ReAct. The external resource must actually be accessible to the system. An appropriate task instruction could therefore be formulated as follows:

> Using the connected German dictionary or corpus tool, search for authentic uses of the expressions "nachhaltige Mobilität," "öffentliche Verkehrsmittel," and "Fahrradnutzung." Examine the retrieved examples and identify frequent or pedagogically useful collocations. Based on these observations, generate brief definitions and example sentences that are comprehensible for B1-level learners.

Within foreign language education, ReAct may therefore be useful when a task requires genuine interaction with dictionaries, corpora, search tools, databases, or other external resources. However, access to such tools does not eliminate the need for teacher supervision; retrieved information and generated examples must still be evaluated for linguistic accuracy, proficiency-level appropriateness, authenticity, and pedagogical relevance.

Stage 5: Model-Assisted Chain-of-Verification Followed by Mandatory Teacher Review

The final stage focuses on the systematic verification of the generated instructional material. Within the proposed framework, Chain-of-Verification (CoVe) functions as a **model-assisted verification mechanism** rather than as a definitive guarantee of accuracy. The model is prompted to formulate and answer verification questions concerning its own output before producing a revised version.

1. Verify the activity you have generated below.
2. Is every factual claim supported by the provided or retrieved sources?
3. Has any unsupported information been introduced?
4. Are the vocabulary and structural complexity appropriate for CEFR B1?
5. Are the comprehension questions consistent with the reading passage?
6. Is the answer key accurate? Answer each verification question separately, identify any problems, revise the activity accordingly, and then provide the corrected version.

CoVe provides an additional opportunity to identify inconsistencies, unsupported claims, answer-key errors, or proficiency-level mismatches. However, model-based self-verification cannot substitute for professional pedagogical judgment. Therefore, mandatory teacher review constitutes the final validation step in the proposed decision tree. The teacher remains responsible for evaluating linguistic accuracy, source alignment, CEFR appropriateness, cultural validity, and the overall instructional suitability of the generated material. This emphasis also preserves the original study's position that ultimate pedagogical accountability remains with the educator.

Within the present scenario, the decision tree operates as a sequence of functionally differentiated processes. Because the teacher directly provides the CEFR outcomes, textbook passage, and supplementary materials, the workflow begins with source-grounded/document-grounded prompting rather than RAG. Because some of these materials are excessively detailed or partially irrelevant, the pedagogically adapted logic of Chain-of-Note is subsequently used to identify relevant information and reduce informational noise. Where the task requires explicit organization of facts, concepts, and relationships, Chain-of-Knowledge is employed to construct structured knowledge evidence that can support subsequent reasoning and generation. If authentic lexical information must be obtained through an actually accessible dictionary, corpus, search system, or API,

ReAct provides a reasoning-and-action workflow for interacting with those external tools. By contrast, when relevant information is not directly supplied by the teacher but must be retrieved from an external indexed knowledge source, Retrieval-Augmented Generation constitutes the appropriate retrieval-based approach. Finally, the generated instructional material undergoes CoVe-assisted verification followed by mandatory teacher review, ensuring that model-based checking remains subordinate to professional pedagogical validation.

Conclusion

Rather than constituting mutually exclusive alternatives, the approaches represented in the decision tree support different informational and pedagogical requirements within the material-development process. Source-grounded or document-grounded prompting is appropriate when teachers directly provide the materials on which generation should be based, whereas Retrieval-Augmented Generation (RAG) becomes relevant when task-related information must be retrieved from an external indexed knowledge source. Maintaining this distinction prevents direct document provision from being conflated with retrieval-based generation.

Within source-rich environments, the pedagogically adapted use of Chain-of-Note can support the filtering of noisy, excessive, or instructionally irrelevant information. Chain-of-Knowledge, by contrast, can be employed when relevant information needs to be organized as structured knowledge evidence before reasoning or content generation. ReAct becomes relevant when the instructional task requires the model to reason while actually interacting with external tools such as dictionaries, corpora, search systems, databases, or APIs. Finally, Chain-of-Verification provides a model-assisted verification stage through which generated content can be systematically questioned and revised; however, the final determination of linguistic accuracy, source fidelity, proficiency-level appropriateness, cultural validity, and pedagogical usability remains the responsibility of the teacher. Thus, the decision tree positions prompt engineering techniques and related generation approaches as complementary decision-support mechanisms rather than replacements for professional pedagogical judgment.

Declaration on the use of artificial intelligence

ChatGPT (GPT-5.6 Sol) and Gemini were employed exclusively for language editing and visualization. After using these tools, the author

thoroughly reviewed and edited the output and takes full responsibility for the final content of the publication.

References

Abaidoo, A. (2026). A Systematic Review of Prompt Engineering for Responsible Use of GenAI in K-12 Schools. *Asian Journal of Research in Computer Science.* https://doi.org/10.9734/ajrcos/2026/v19i1812

Akhmadullina, R., Valiakhmetova, N., & Yarmakeev, I. (2026). Methodological aspects of developing promptengineering competence to solve pedagogical problems in the structure of future teacher's media competence. *Philology and Culture.* https://doi.org/10.26907/2782-4756-2026-83-1-233-240

Alyahyan, E. A., Ada, M. B., & Lever, J. (2025). WIP: A Pedagogical Prompt Engineering Framework for LLM-Based Feedback in Higher Education (PPE-LLM). *2025 IEEE Frontiers in Education Conference (FIE)*, 1-5. https://doi.org/10.1109/fie63693.2025.11328552.

Antonopoulou, H., Matzavinou, P., Giannoukou, I., & Halkiopoulos, C. (2025). Teachers' Digital Leadership and Competencies in Primary Education: A Cross-Sectional Behavioral Study. *Education Sciences*. https://doi.org/10.3390/educsci15020215.

Aspandi, A., & Muttaqin, M. (2025). Transforming Teacher Roles in Indonesia's Digital Era: Enhancing Learning Effectiveness and Student Engagement. *Journal of General Education and Humanities*. https://doi.org/10.58421/gehu.v4i4.616

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32(4)*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Caena, F., & Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (DigCompEdu). *European Journal of Education, 54*(3), 356–369. https://doi.org/10.1111/ejed.12345

Cain, W. (2023). Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends, 68,* 47 - 57. https://doi.org/10.1007/s11528-023-00896-0

Chen, K. (2025). Teacher Role Reshaping in Digital Transformation of Education: Innovation of Teaching Model Based on Artificial Intelligence. *Lecture Notes in Education Psychology and Public Media*. https://doi.org/10.54254/2753-7048/2026.ht28860

Chiu, T., Ahmad, Z., Ismailov, M., & Sanusi, I. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100171

Dourvas, I., Kokkonis, G., & Kontogiannis, S. (2026). Reconceptualizing Prompt Engineering as Reflective Professional Practice: A Framework for Teacher Development. *Electronics.* https://doi.org/10.3390/electronics15050930

Elsayary, A., Kuhail, M. A., & Hojeij, Z. (2025). Examining the Role of Prompt Engineering in Utilizing Generative AI Tools for Lesson Planning: Insights From Teachers' Experiences and Perceptions. *Human Behavior and Emerging Technologies.* https://doi.org/10.1155/hbe2/9986139

European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. Publications Office of the European Union. https://doi.org/10.2766/153756

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union, L 2024/1689*.

Galindo-Domínguez, H., Delgado, N., Campo, L., & Losada, D. (2024). Relationship between teachers' digital competence and attitudes towards artificial intelligence in education. *International Journal of Educational Research*. https://doi.org/10.1016/j.ijer.2024.102381

Huang, T., Wu, C., & Wu, M. (2025). Developing pre-service language teachers' GenAI literacy: An interventional study in an English language teacher education course. *Discover Artificial Intelligence, 5*, Article 163. https://doi.org/10.1007/s44163-025-00435-1

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kalimullina, O., Tarman, B., & Stepanova, I. (2020). Education in the Context of Digitalization and Culture: Evolution of the Teacher's Role, Pre-pandemic Overview. *Journal of Ethnic and Cultural Studies*. https://doi.org/10.29333/ejecs/629

Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals, 51*(1), 205–218. https://doi.org/10.1111/flan.12318

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: a systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*. https://doi.org/10.1186/s41239-025-00503-7

Leung, C.-H. (2024). Promoting optimal learning with ChatGPT: A comprehensive exploration of prompt engineering in education. *Asian Journal of Contemporary Education.* https://doi.org/10.55493/5052.v8i2.5101

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship.* https://doi.org/10.1016/j.acalib.2023.102720

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Markauskaite, L., Carvalho, L., & Fawns, T. (2023). The role of teachers in a sustainable university: from digital competencies to postdigital capabilities. *Educational Technology Research and Development, 71*, 181 - 198. https://doi.org/10.1007/s11423-023-10199-z

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO.

Mikeladze, T., Meijer, P., & Verhoeff, R. (2024). A comprehensive exploration of artificial intelligence competence frameworks for educators: A critical review. *European Journal of Education*. https://doi.org/10.1111/ejed.12663

Moguš, A. M. (2025). Using prompt engineering to foster critical thinking in higher education. *ELEARNING2025*. https://doi.org/10.46793/elearning2025.235mm

Ng, D., Leung, J., Su, J., Ng, R., & Chu, S. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development, 71*, 137 - 161. https://doi.org/10.1007/s11423-023-10203-6

Oreški, D., Kišić, A., & Rožman, M. (2025). Prompt Engineering as an AI Literacy Competence: A Framework for Learners and Educators. *ELEARNING2025*. https://doi.org/10.46793/elearning2025.160o

Qian, Y. (2025). Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications. *Journal of Educational Computing Research*, *63*(7-8), 1782-1818.

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (Y. Punie, Ed.). Publications Office of the European Union. https://doi.org/10.2760/159770

Reinius, H., Kaukinen, I., Korhonen, T., Juuti, K., & Hakkarainen, K. (2022). Teachers as transformative agents in changing school culture. *Teaching and Teacher Education.* https://doi.org/10.1016/j.tate.2022.103888

Ren, X., & Wu, M. (2025). Examining Teaching Competencies and Challenges While Integrating Artificial Intelligence in Higher Education. *TechTrends*, *69,* 519 - 538. https://doi.org/10.1007/s11528-025-01055-3

Rey, K. (2025). Harnessing AI in secondary education in Chinese Hong Kong: the role of prompt engineering. *Journal of China Computer-Assisted Language Learning, 0*. https://doi.org/10.1515/jccall-2025-0016

Saefurrohman, Mariana, N., Hartono, B., Candra Rina, S. N. & Lusiana, V. (2026). Strengthening Vocational High School Teachers' Pedagogical Competence in Developing AI-Based Learning Materials through Prompt Engineering Training. *Community Service Journal: Adarma Acitya*. https://doi.org/10.26877/adarmaacitya.v2i1.572

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2025). *A systematic survey of prompt engineering in large language models: Techniques and applications*. arXiv. https://arxiv.org/abs/2402.07927

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences, 12*(8), Article 569. https://doi.org/10.3390/educsci12080569

Serra, P., & Oliveira, Â. (2025). AI-Powered Prompt Engineering for Education 4.0: Transforming Digital Resources into Engaging Learning Experiences. *Education Sciences*. https://doi.org/10.3390/educsci15121640

Schmitt, C., & Brutzer, A. (2025). Generative artificial intelligence in vocational education and training: A framework for sustainable teacher competence development. *Vocation, Technology & Education*. https://doi.org/10.54844/vte.2025.0940

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in Teacher Education: A scoping review. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100169

Tenberga, I., & Daniela, L. (2024). Artificial Intelligence Literacy Competencies for Teachers Through Self-Assessment Tools. *Sustainability*. https://doi.org/10.3390/su162310386

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.

Urban, M., Lukavský, J., Brom, C., Hein, V., Svacha, F., Děchtěrenko, F., & Urban, K. (2025). Prompting for creative problem-solving: A process-mining study. *Learning and Instruction, 99*, 102156. https://doi.org/10.1016/j.learninstruc.2025.102156

Velander, J., Taiye, M. A., Otero, N., & Milrad, M. (2024). Artificial Intelligence in K-12 Education: eliciting and reflecting on Swedish teachers' understanding of AI and its implications for teaching & learning. *Education and Information Technologies, 29*, 4085 - 4105. https://doi.org/10.1007/s10639-023-11990-4

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*, 1-29. https://doi.org/10.1186/s41239-024-00448-3

Wohlfart, O., & Wagner, I. (2023). Teachers' role in digitalizing education: an umbrella review. *Educational Technology Research and Development, 71*, 339 - 365. https://doi.org/10.1007/s11423-022-10166-0

Yue, M., Jong, M., & Ng, D. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Education and Information Technologies, 29*, 19505 - 19536. https://doi.org/10.1007/s10639-024-12621-2

Zou, D., Xie, H., & Kohnke, L. (2025). Navigating the Future: Establishing a Framework for Educators' Pedagogic Artificial Intelligence Competence. *European Journal of Education*. https://doi.org/10.1111/ejed.70117

Part III
Materials and Task Design

A Stage-Oriented Model of AI-Personalized German Language Preparation from A1 to B2

Betim Neyran Can [1]

Abstract
Artificial intelligence is no longer a peripheral concern in language education; for many learners, it has already become part of the ordinary landscape of studying, practising, and revising a new language. The more difficult question is how that role ought to evolve as learners move from dependence on basic linguistic support toward greater communicative and critical independence. This chapter proposes a stage-oriented account of AI-supported personalization in German preparatory programmes spanning A1 to B2. Instead of equating personalization with an ever-increasing amount of automated assistance, the model links the type and intensity of support to proficiency, task purpose, learner independence, and teacher judgment. At A1–A2, AI is most useful when it stays close to the curriculum: short vocabulary work, pronunciation rehearsal, controlled grammar practice, and predictable exchanges. At B1–B2, its role can widen to include comparison, revision, mediation, register awareness, argument development, and reflective feedback. At every stage, however, pedagogical responsibility continues to rest with the teacher, particularly in calibrating difficulty, checking the quality of generated language, protecting meaningful human interaction, and helping learners judge instead of simply accept AI output. The chapter closes by translating these principles into a practical framework in which technology remains subordinate to pedagogical purpose.
Keywords: artificial intelligence; German as a foreign language; personalized learning; CEFR; generative AI

Introduction

Generative AI has changed the practical conditions under which language learning now takes place. A student who is unsure about a word order pattern can ask for more examples immediately; another can rehearse a dialogue before speaking in class; a third can compare two versions of a paragraph and ask why one sounds more natural. Their appeal lies partly in their immediacy: support can appear at the exact moment a learner encounters difficulty. Yet immediacy should not be confused with educational value. Having more language available is not, by itself, the same as learning more effectively.

[1] Lecturer, Istanbul University-Cerrahpaşa, School of Foreign Languages, Department of Foreign Languages, English Preparatory Program, ORCID: 0000-0001-5626-4492, *betim.can@iuc.edu.tr*, Türkiye

This distinction becomes especially consequential in German preparatory programmes. An A1 learner who is still stabilizing articles, verb position, and high-frequency vocabulary needs a very different kind of support from a B2 learner working on argumentation, register, and extended discourse. The same AI response may be helpful for one learner, distracting for another, and far beyond the independent competence of a third. The relevant question, then, is not how extensively a programme uses AI, but whether that use is developmentally appropriate for what learners can reasonably understand, attempt, and evaluate.

It is from this problem of developmental fit that the stage-oriented model proposed here emerges. The model deliberately begins with learning instead of with a catalogue of available tools. It is the developmental movement from A1 to B2 and the changing demands this places on learners: first building a reliable linguistic base, then managing longer stretches of language, making choices, interpreting context, and evaluating alternatives. AI is assigned a different role as these demands change. At the lower levels it mainly supports practice and scaffolding; at the higher levels it becomes more useful as a partner for comparison, rehearsal, revision, and reflection.

The position taken throughout is deliberately human-centred. Speed, scale, and endless variation are genuine advantages of AI, but they do not resolve the educational questions that matter most. Generated language may be fluent and still be inaccurate, culturally flat, pragmatically awkward, or simply too polished to be useful for learning. A learner can likewise receive so much assistance that the difficult but productive part of the task disappears. Teacher mediation is therefore not an extra layer added after the technology; it is what gives the technology pedagogical direction (Harsch et al., 2025; Hartmann, 2024).

Three closely connected aims structure the discussion. First, it clarifies what meaningful personalization can mean in German language preparation from A1 to B2. Second, it translates this understanding into a practical stage-oriented framework for vocabulary, grammar, pronunciation, reading, writing, speaking, mediation, and self-directed learning. Third, it identifies the conditions under which AI use can strengthen rather than weaken learner agency, classroom interaction, and pedagogical coherence. The focus is conceptual and practice-oriented: the purpose is not to claim that AI automatically improves outcomes, but to offer a principled way of deciding when, why, and how it may add educational value.

From Technological Adaptation to Pedagogical Personalization

Personalized learning is commonly understood as the deliberate adjustment of instruction in response to differences in learners' knowledge, aims, interests, preferences, and pace. In language education, this may involve different amounts of vocabulary support, alternative explanations of grammar, flexible task difficulty, targeted pronunciation practice, or materials connected to learners' academic and professional purposes. Generative AI appears to extend these possibilities because it can respond immediately to a learner's request and produce multiple versions of the same content. But responsiveness at the level of output should not be mistaken for pedagogically meaningful personalization.

A system can simplify a text without knowing whether the simplification removes the discourse features a learner needs to notice. It can correct a sentence without knowing whether the learner understands the correction. It can produce an apparently B1-level dialogue whose topic, cultural assumptions, or interactional demands are actually too difficult. Conversely, it can over-simplify language for an intermediate learner and unintentionally restrict development. These examples illustrate the difference between surface-level adaptation and deeper instructional alignment. The latter requires attention to what the learner is ready to do, what the task is intended to develop, how much support is productive, and what should remain the learner's own cognitive work.

From this perspective, personalization is better understood as a sequence of pedagogical decisions than as a built-in property of a digital system. The first question is not 'What can the AI do?' but 'What does the learner need to learn next?' The second is 'What kind of assistance will make that learning more likely without replacing the learning process itself?' Only then does the question of technology become relevant. Such a needs-first orientation matters especially in language education, where progress is not simply an accumulation of correct forms. Learners must gradually develop control over meaning, interaction, genre, register, discourse, and cultural context.

This distinction becomes especially useful when the same AI function may be productive in one situation and counterproductive in another. Immediate correction can be useful when an A1 learner is practising a tightly defined verb pattern. The same correction function may be insufficient for a B2 argumentative text, where the most important issues concern reasoning, organization, rhetorical purpose, and audience. Similarly, a generated model answer may reduce anxiety before a beginner role-play, but a polished model essay can deprive an advanced learner of planning, drafting,

and revision. Pedagogically defensible personalization thus depends on calibrated, not maximal, assistance.

A stage-oriented perspective makes this calibration more concrete. It links the degree of AI support to learners' growing ability to interpret feedback, tolerate ambiguity, make linguistic choices, and reflect on those choices. This is consistent with the broader literature positioning learners and teachers as active agents who transform technology-supported activities rather than passively receiving automated instruction (Yang & Kyun, 2022). Personalization becomes a relationship among learner readiness, task design, teacher judgment, and technological affordances.

AI in Language Learning: Affordances and Boundaries

The current landscape of AI-supported language learning includes a strikingly broad set of functions: speech recognition, automated writing support, adaptive vocabulary practice, conversational simulation, intelligent tutoring, text generation, translation, summarization, and feedback. These tools can extend practice beyond scheduled classroom time and make additional examples or explanations available when a teacher is not immediately present. In German as a foreign language, large language models can likewise help structure self-study and connect independent practice with classroom work.

For classroom teachers, one of the most tangible benefits is the possibility of differentiation. A teacher can use AI to create several versions of a reading text, generate glossaries, adjust question difficulty, or produce follow-up tasks that require different degrees of support. Such uses can reduce preparation time in heterogeneous classes and make it easier to preserve a shared topic while varying the linguistic demands placed on individual learners. Differentiation is most valuable when it keeps learners within a common curricular and social environment rather than sending each student down an isolated pathway.

Formative feedback provides another productive, though more complicated, area of use. AI often identifies recurring grammatical, lexical, spelling, and syntactic problems and provide rapid reformulations. This immediacy can be motivating and can help learners notice patterns in their own production. Yet automated feedback is uneven. It may identify a grammatically possible expression without recognizing that it is pragmatically odd, too formal, culturally narrow, or inconsistent with the intended meaning. It may also privilege sentence-level correctness over coherence, argumentation, and rhetorical effectiveness (Renandya & Floris, 2024).

Feedback is therefore useful only when learners can understand it and when it remains connected to an explicit learning objective.

Rehearsal can be equally valuable, especially for learners who benefit from a private first attempt before speaking in front of others. Conversational systems can simulate everyday situations, interviews, complaints, negotiations, or discussions. For learners who are hesitant to speak publicly, such rehearsal may lower the threshold for participation and allow repetition without embarrassment. It can also provide additional practice when access to proficient speakers is limited. Still, a simulated conversation is not equivalent to interpersonal communication. AI does not bring lived experience, genuine intentions, emotional accountability, or the unpredictability of a human interlocutor. Its best use is often preparatory: learners rehearse with AI and then transfer the language to pair work, role-play, debate, or collaborative problem-solving.

With increasing proficiency, mediation and text transformation create a qualitatively different space for learning. Learners may compare an original text with an AI-generated summary, identify what has been omitted, adapt information for a different audience, or evaluate whether a culturally specific meaning has been preserved. These activities are particularly promising because they make AI output an object of language work instead of a ready-made answer. Hartmann similarly highlights the potential of AI applications to support mediation in German as a foreign language. The pedagogical gain lies not in accepting the generated version, but in examining and transforming it.

These affordances, however, come with clear limits. AI may hallucinate facts, generate unnatural collocations, reproduce bias, or present uncertain claims with confidence. It can likewise encourage dependency if learners habitually ask it to translate, formulate, or rewrite before attempting a task themselves. Responsible integration therefore requires an explicit division of labor: AI can expand opportunities for practice and comparison, but learners remain responsible for meaning and teachers remain responsible for pedagogical quality, cultural context, and ethical oversight.

The Stage-Oriented Model

The model developed here brings together three interdependent dimensions of AI-supported German preparation: learner proficiency, instructional purpose, and degree of teacher mediation. CEFR levels provide a developmental reference point, but they are not treated as rigid boxes. Learners at the same nominal level may differ in vocabulary size, pronunciation, confidence, literacy, prior schooling, and ability to work

independently. The model thus uses CEFR stages as a planning framework while preserving room for individual variation.

The first principle is developmental alignment. AI-supported tasks should match the linguistic and cognitive demands learners can reasonably manage. At A1 and A2, this usually means short input, familiar topics, limited lexical range, predictable interaction, and explicit models. At B1 and B2, tasks can involve longer texts, multiple perspectives, more open-ended interaction, register choices, implicit meaning, and independent revision. The second principle is purposeful assistance. Every use of AI should have a named function: retrieval practice, pronunciation rehearsal, grammar reinforcement, idea organization, cohesion review, register comparison, or another clearly defined purpose. Unrestricted use makes it difficult to know whether technology is supporting learning or simply completing work.

Learner autonomy, similarly, should be developed progressively instead of taken for granted. Beginners need stronger external control because they often lack the linguistic knowledge required to judge AI output. As proficiency develops, learners can be given more responsibility for prompting, evaluating, rejecting, and revising generated suggestions. This does not mean teacher mediation steadily disappears. Rather, it changes form. At A1, mediation may involve pre-selecting vocabulary and checking every model sentence. At B2, it may involve asking learners to defend why they accepted one reformulation and rejected another.

Equally important is a deliberate return to human interaction. Individual AI-supported practice should return to human communication. A vocabulary drill should feed into a pair exchange; an AI-supported reading should lead to discussion; a revised paragraph should be peer-reviewed; a simulated negotiation should be performed with classmates. This principle protects the social nature of language learning and prevents personalization from becoming isolation. It also makes the classroom the place where language is tested against real interlocutors, competing interpretations, and unpredictable responses.

The final principle is reflective verification. Learners should gradually learn that fluent AI output is not self-validating. They need routines for checking accuracy, appropriateness, evidence, and register. At lower levels, verification may be as simple as comparing a correction with a teacher-approved model. At higher levels, it can involve consulting dictionaries and course materials, checking claims against reliable sources, comparing alternative formulations, and explaining the pragmatic consequences of different choices. In this way, AI literacy develops alongside language proficiency rather than as a separate technical skill.

A1–A2: AI as Structured Scaffolding

At A1 and A2, the central task is still foundational: learners are trying to make a relatively small linguistic repertoire dependable enough for everyday use. High-frequency vocabulary must become retrievable; basic sentence patterns must become familiar; pronunciation needs repeated practice; and routine communicative exchanges need to become manageable. At this point, AI is most useful not because it is creative, but because it can make carefully bounded practice repeatedly available. It can provide many short opportunities to repeat, retrieve, listen, notice, and try again at an individual pace.

Vocabulary work illustrates the point particularly well. Rather than asking an AI system to 'teach German vocabulary,' the teacher can specify a small lexical set connected to the current unit and ask for controlled retrieval tasks. One learner may need extra practice with family members, another with classroom objects, and another with time expressions. The content remains aligned with shared curricular goals, but the amount and focus of practice vary. AI can generate matching items, cloze sentences, simple categorization tasks, or short question-and-answer exchanges. The teacher still needs to verify the lexical choices and prevent the system from introducing unnecessary synonyms or low-frequency items.

Grammar practice benefits from the same kind of boundary-setting. A1 learners may practise verb position, present-tense conjugation, articles, negation, or simple accusative structures using teacher-approved sentence frames. A2 learners can gradually work with modal verbs, separable verbs, common past-tense forms, subordinate clauses, and routine comparisons. What matters here is the possibility of targeted repetition. If a learner repeatedly misplaces the verb, the system can generate a small set of sentence-building tasks focused on that pattern. The task should remain short enough for the learner to notice the form rather than become overwhelmed by new vocabulary.

Pronunciation is another area in which individualized rehearsal can be genuinely useful. Learners may rehearse short words and phrases, compare their production with a model, and repeat difficult sound combinations. Speech-recognition systems can provide additional feedback, but their judgments should not be treated as definitive. Accent variation, microphone quality, and system limitations can affect results. The teacher still needs to identify which pronunciation features are communicatively important and to prevent learners from interpreting a software score as an absolute measure of intelligibility.

Short conversational rehearsals work best when the communicative purpose is familiar and the interaction remains reasonably predictable. A1 learners can practise introducing themselves, ordering food, asking for a price, or giving a simple direction. A2 learners can handle slightly longer exchanges involving appointments, travel problems, invitations, past experiences, or routine complaints. The system can be asked to keep its turns short, use known vocabulary, and avoid suddenly increasing complexity. After rehearsal, learners should perform a related exchange with a peer, where timing, misunderstanding, and spontaneous response become part of the task.

In classroom practice, this can produce a straightforward but pedagogically purposeful A1–A2 sequence. First, the teacher introduces the target language and provides a reliable model. Second, learners complete a narrowly defined AI-supported practice activity at a pace suited to their needs. Third, they compare their responses with a teacher-approved reference and mark anything they do not understand. Fourth, the class returns to pair or group communication, where the target language is used for a meaningful purpose. AI thus increases the quantity of practice without becoming the source of curricular direction.

The principal risk at beginner levels is epistemic vulnerability: beginners often cannot tell when an AI response is inaccurate, unnatural, or too advanced. They may internalize a plausible-looking form simply because it is presented fluently. For that reason, beginner prompts should be constrained, generated materials should be checked, and explanations should use language learners can understand. Personalization at A1 and A2 should primarily adjust pace, repetition, focus, and the amount of scaffolding—not replace common learning outcomes or teacher guidance.

B1–B2: AI as a Resource for Reflection and Communicative Expansion

By B1 and B2, the pedagogical value of AI begins to shift. Learners are no longer working only toward accurate isolated sentences; they must sustain interaction, organize longer texts, explain viewpoints, interpret intentions, adapt register, and connect ideas coherently. They also possess more linguistic resources with which to question the output they receive. At this point, AI can move from tightly controlled scaffolding toward guided comparison, revision, rehearsal, and reflection.

Reading makes this shift especially visible. At B1, learners can work with adapted or authentic short texts such as announcements, interviews, news items, blogs, or informational webpages. AI may generate a glossary,

identify potentially difficult expressions, or create comprehension questions, but learners should first attempt the text independently. Only afterward should they consult AI for clarification. This sequence preserves inferencing and reading strategies. At B2, learners can compare perspectives, identify implicit claims, examine tone, or ask how the same content might be framed for different audiences. AI-generated interpretations should remain provisional and be checked against the source text and classroom discussion.

Writing support should undergo a similar change: the emphasis can move away from simple correction and toward inquiry. At B1, AI may help learners examine paragraph organization, topic sentences, cohesion, repeated vocabulary, and common grammatical patterns. The learner should produce the first draft. The system may then identify unclear transitions or ask questions about missing information, but it should not simply replace the draft with a polished version. At B2, learners can use AI to test the clarity of a thesis, compare alternative structures, explore register, or receive questions that expose weaknesses in reasoning. The most productive prompt is often not 'Rewrite this better' but 'Which parts of my argument are unclear, and what questions should I answer before revising?'

This distinction is important because polished output can create the appearance of development without demonstrating corresponding learner control. When AI produces an elegant paragraph, the text may exceed the learner's independent repertoire. The learner may submit a stronger product while developing little control over the language that created it. A learning-oriented approach therefore requires visible decision-making. Students should identify which suggestions they accept, which they reject, and why. They can compare first and final drafts and briefly explain how particular revisions improved precision, cohesion, or appropriateness. Such procedures turn feedback into metalinguistic reflection (Joo, 2024; Zeevy-Solovey, 2024).

Speaking work at B1 and B2 can afford considerably more unpredictability. Learners may rehearse agreement and disagreement, complaints, advice, interviews, negotiations, or discussion of social issues. At B2, simulations can include ambiguity, follow-up questions, changes of topic, or competing viewpoints. Yet the endpoint should still be human interaction. A learner might use AI to prepare a position for a debate, but the debate itself should require listening to classmates, responding to unexpected arguments, and adjusting language in real time. These are precisely the social and strategic dimensions that automated interaction cannot fully reproduce.

Mediation and register awareness become particularly fertile areas for work at these levels. Learners may ask AI to reformulate the same message for a friend, a university office, a workplace, or an academic audience and then compare the results. They can identify where politeness, directness, lexical choice, or sentence structure changes and discuss whether the generated version is actually appropriate. Similarly, learners can summarize a German text for a different audience, compare their version with an AI-generated one, and evaluate what information was lost or distorted. Such tasks make cultural and pragmatic judgment visible rather than treating language as a set of interchangeable sentences.

The learning sequence at B1–B2 can therefore become more open-ended: independent attempt, narrowly defined AI consultation, evaluation of the output, learner-led revision, and human interaction or reflection. The independent attempt is crucial because it establishes what the learner can do without assistance. The AI consultation then becomes diagnostic and comparative instead of substitutive. Evaluation requires the learner to judge accuracy, relevance, register, and usefulness. Revision returns responsibility to the learner, and the final human stage connects individualized support with the social purposes of language.

As learners become more independent, personalization can likewise respond more directly to their academic, professional, and personal purposes. A student preparing for academic study may need formal email writing and seminar discussion; another may need workplace interaction; another may be preparing for a proficiency examination. AI may help generate context-specific practice while the teacher ensures that the tasks remain linguistically appropriate and connected to the program's outcomes. Personalization therefore becomes richer as proficiency grows, but it should also become more demanding: greater technological freedom must be accompanied by greater critical responsibility.

Teacher Mediation in Mixed-Level German Preparatory Programs

Few contexts make the practical case for AI-supported differentiation as clearly as the mixed-level classroom. Teachers often need to maintain a common syllabus while responding to learners who differ in speed, confidence, prior exposure, and skill profiles. AI may help vary the amount of support without requiring entirely separate lesson plans. The key is to preserve shared communicative goals. Learners may discuss the same topic, solve the same problem, or work toward the same final task while receiving different forms of preparation.

For instance, a common lesson on sustainable travel might provide A2 learners with a short text, a vocabulary bank, and sentence starters, while B1 learners work with a longer text and more open questions. In a B1–B2 group, one set of learners might receive guiding questions and a cohesion checklist, while more advanced learners compare competing viewpoints and evaluate register. AI can generate these variants quickly, but the teacher decides what counts as an appropriate difference. The purpose is not to individualize everything; it is to remove unnecessary barriers while keeping learners connected to one another.

Flexible group roles can extend this inclusive logic. In group tasks, learners can act as summarizer, questioner, language monitor, evidence checker, or discussion leader. AI can help prepare a learner for a particular role, but the role is performed within a human group. This arrangement makes differentiation socially productive. Learners contribute at different levels of complexity while still participating in a common communicative event.

Teacher mediation remains indispensable because automatically generated differentiation is not reliably well calibrated. A supposedly easier version may contain unfamiliar idioms; a generated B2 task may be cognitively simple despite complex vocabulary; a cultural example may be inappropriate for the learners' context. Teachers thus need to evaluate generated materials not only for grammatical correctness but also for conceptual difficulty, discourse demands, cultural assumptions, and curricular fit. This is pedagogical work, not merely quality control.

A useful way to describe the teacher's role here is orchestration. The teacher decides when AI should be used, how much assistance it should provide, when learners should work independently, and when the class should return to shared interaction. This orchestration prevents the classroom from fragmenting into a collection of private learner-machine exchanges. It also ensures that technology serves the learning sequence instead of determining it. As research on AI-supported language education repeatedly suggests, effective integration depends on the relationship among tools, tasks, teachers, and learners rather than on the technology in isolation.

Learner Autonomy, Prompt Literacy, and Productive Struggle

AI is frequently linked with learner autonomy, largely because explanations, examples, and additional practice can be accessed on demand. Learners may decide when to ask for help, how many examples to request, and which difficulty to revisit. This flexibility can strengthen self-regulation, particularly when learners are taught to set goals, monitor recurring

errors, and reflect on progress. Yet access to instant assistance can also produce the opposite effect. If learners routinely outsource planning, translation, formulation, and revision, they may become less willing to tolerate uncertainty and less capable of independent performance.

For this reason, the boundary between assistance and substitution is central to the model. Productive language learning involves moments of difficulty: retrieving a word, deciding between structures, organizing an argument, or trying to understand a text before reaching for a translation. These moments are not inefficiencies to be eliminated. They are often the points at which noticing, hypothesis testing, and self-regulation occur. A responsible AI pedagogy should preserve enough productive struggle for learning to remain cognitively active.

A practical principle follows from this: attempt before assistance. Learners first produce a sentence, interpretation, outline, or draft. Only then do they consult AI for a specified purpose. A second rule is 'explain the change.' If learners accept a correction or reformulation, they should be able to state what changed and why. A third is 'verify when stakes are high.' Claims, unfamiliar expressions, and culturally sensitive formulations should be checked against reliable sources, dictionaries, course materials, or teacher guidance. These routines make AI use visible as part of learning rather than an invisible shortcut.

Prompt literacy forms one part of this developing autonomy. Effective prompts require learners to articulate the task, level, purpose, constraints, and kind of feedback they need. In this sense, prompting can become a metacommunicative competence: learners must think about what they are trying to do with language before they can ask the system for useful assistance. At the same time, prompt skill should not be romanticized as a purely technical competence. A well-written prompt cannot guarantee an accurate or pedagogically appropriate response. Critical evaluation remains more important than prompt elegance.

Responsibility for prompting can therefore be transferred gradually as learners progress from A1 toward B2. Beginners may work with teacher-provided prompts such as 'Ask me five A1 questions about my daily routine. Use only present tense and short sentences.' Intermediate learners can modify prompts, specify the type of feedback they want, and compare different outputs. By B2, learners can design prompts that ask for counterarguments, register comparisons, or diagnostic questions instead of finished products. The development of autonomy is therefore not a withdrawal of guidance but a gradual transfer of informed decision-making.

Risks, Ethics, and Conditions for Responsible Use

A credible model of AI-supported personalization has to begin from a simple fact: generative systems are fallible. They can produce factual inaccuracies, invented references, unnatural language, and confident explanations that are incomplete or wrong. In language education, these errors are especially consequential because learners may not possess the knowledge needed to detect them. Beginners are vulnerable to internalizing incorrect forms; more advanced learners may be misled by subtle problems of register, collocation, or pragmatics. Teacher verification and learner checking routines are thus core pedagogical requirements, not optional safeguards.

Questions of bias and cultural representation are no less important. AI systems may reproduce dominant linguistic norms or present one variety of German as universally appropriate. They may flatten regional, social, or intercultural variation and generate stereotyped examples. A human-centered approach treats such output critically. Teachers may sometimes turn problematic responses into learning material by asking students to identify assumptions, compare alternatives, or discuss why a formulation may be acceptable in one context but not another. This transforms an AI limitation into an opportunity for sociolinguistic awareness, provided that the teacher can safely and accurately frame the discussion.

Privacy introduces a further institutional and ethical concern. AI-supported platforms may collect writing samples, speech recordings, interaction histories, and other learner data. Institutions should consider what information is collected, where it is stored, who can access it, and whether learners understand the implications of use. Sensitive personal information should not be required for routine language practice. Responsible implementation also means choosing tools and procedures that are consistent with institutional data-protection requirements (Mohamed, 2024; Häußler et al., 2026).

Academic integrity and assessment, meanwhile, require policies that are explicit enough for both teachers and learners to understand. If AI is permitted for brainstorming, language checking, or formative feedback, learners should know what must be acknowledged and what remains their own responsibility. If a task is intended to measure independent writing or speaking, unrestricted AI assistance can undermine validity. Assessment design may therefore need to include process evidence, in-class production, oral follow-up, reflection on revisions, or documented use of AI. The aim is not merely to detect AI use but to align assessment with the competence the program intends to measure (Joo, 2024).

A less visible, but equally significant, risk is that social learning can gradually be weakened. Language is learned through relationships, shared attention, negotiation of meaning, repair, emotion, and participation in communities. AI can simulate aspects of dialogue, but it does not eliminate the need for genuine interaction. If personalization leads learners to spend more time in private exchanges with a system and less time responding to peers, an apparent increase in individualized practice may come at the cost of communicative development. The model proposed here thus treats human interaction as non-negotiable. AI prepares, extends, or supports communication; it does not become the primary social environment of language learning.

In practical terms, responsible integration can be anchored in five conditions: developmental appropriateness, transparency of purpose, verification of output, preservation of learner agency, and reintegration into human communication. These conditions shift attention away from whether AI is inherently 'good' or 'bad' and toward the quality of the pedagogical decisions surrounding it. They also provide institutions with a practical basis for teacher training and local policy.

Implications for Curriculum, Assessment, and Teacher Development

These principles also have direct implications for curriculum design. AI activities should not appear as isolated digital extras but should be mapped onto existing learning outcomes. At A1 and A2, curriculum documents can identify where controlled AI practice may support vocabulary, pronunciation, grammar, and routine interaction. At B1 and B2, they can specify where AI may support planning, revision, mediation, register awareness, or discussion preparation. This mapping helps prevent tool-driven teaching and makes it easier to evaluate whether a particular use actually contributes to progression.

AI literacy should likewise be treated as developmental instead of as a one-off technical orientation. Beginners need simple rules: do not trust every answer, use teacher-approved prompts, and compare corrections with course models. Intermediate learners can learn to request specific kinds of feedback, identify recurring errors, verify unfamiliar language, and document revisions. By B2, learners should be able to question AI-generated claims, evaluate pragmatic appropriateness, compare registers, and explain how AI assistance affected their final product. AI literacy thus becomes intertwined with linguistic and metacognitive development.

Assessment design needs to distinguish clearly between learning with AI and demonstrating competence independently. Formative activities may encourage AI-supported feedback because the purpose is improvement. Summative tasks may require restricted or no AI use when independent performance is the construct being assessed. Between these poles, learning-oriented assessment can ask students to submit an initial draft, a record of selected feedback, a revised version, and a short rationale. Such designs make decision-making visible and reward critical revision instead of invisible dependence.

Sustained teacher development is the condition that makes these curricular ambitions realistic. Training should move beyond demonstrations of popular tools. Teachers need opportunities to evaluate generated materials, calibrate CEFR difficulty, design prompts with clear pedagogical purposes, recognize bias and hallucination, address privacy and academic integrity, and decide when a non-AI activity is preferable. They also need time to share examples of successful and unsuccessful use. Professional judgment develops through comparison, reflection, and adaptation, not through a checklist of applications (Häußler et al., 2026; Renandya & Floris, 2024).

Institutions can make this work more sustainable by establishing transparent guidelines while allowing room for small-scale, evidence-informed experimentation. A sustainable approach does not require every teacher to use AI in every lesson. It requires a shared framework for deciding when the technology adds value and how risks will be managed. This is particularly important in preparatory programs, where curricular coherence and equitable access matter. Students should not be disadvantaged because they have less experience with AI tools, weaker devices, or limited access outside class.

Curriculum planning should also protect spaces that are intentionally and irreducibly human. Seminar discussion, peer feedback, role-play, collaborative writing, oral presentations, teacher conferences, and intercultural reflection are not residual practices left over from a pre-AI classroom. They are central sites of language development. The purpose of AI integration is to make these interactions better prepared, more accessible, or more reflective—not to reduce their importance.

A Practical Implementation Framework

For classroom use, the model can be expressed as a recurring pedagogical cycle: diagnose, design, attempt, assist, evaluate, interact, and reflect. The cycle begins with diagnosis. The teacher identifies a specific learning

need: perhaps article use at A1, word order at A2, paragraph cohesion at B1, or register and argumentation at B2. Diagnosis prevents the technology from becoming the starting point.

The design stage determines whether AI is appropriate and, if so, what it is allowed to do. The teacher defines the target language, expected output, level of complexity, and boundaries of assistance. At lower levels, prompts are likely to be teacher-provided and tightly constrained. At higher levels, learners can participate in prompt design. The attempt stage then requires learners to do something before receiving automated help. This may be a first answer, a short recording, an interpretation, an outline, or a draft.

During the assist stage, AI performs a narrow function. It may generate extra retrieval items, highlight a recurring pattern, ask diagnostic questions, offer alternative formulations, or simulate a brief exchange. The evaluate stage requires the learner to judge the output. Is it correct? Is it understandable? Does it fit the intended register? Does it preserve the intended meaning? At A1, evaluation may rely heavily on a teacher-approved key. At B2, it can become an open comparative task.

The interaction stage returns the language to people. Learners use the prepared material in pair work, discussion, peer review, role-play, presentation, or collaborative problem-solving. Finally, reflection asks what changed. Learners may identify one useful correction, one rejected suggestion, one recurring difficulty, and one goal for the next task. This short reflection is important because it turns a sequence of tool use into a sequence of learning decisions.

The framework is intentionally platform-agnostic. Platforms will change, features will be added, and institutional access will vary. A pedagogical model should remain useful despite those changes. By anchoring decisions in proficiency, purpose, mediation, verification, and human interaction, teachers can evaluate new systems without redesigning their educational philosophy around each technological development.

A stage-oriented perspective challenges two assumptions that recur in discussions of AI and language education. The first is that more personalization is automatically better. In practice, excessive personalization can fragment the curriculum, isolate learners, and make it difficult to sustain common standards. The second is that learner autonomy increases whenever teacher control decreases. In reality, learners may become more dependent when powerful tools provide answers too quickly. Meaningful autonomy requires informed choices, critical evaluation, and responsibility for one’s own language production.

Within this model, scaffolding and autonomy are not opposites; they belong to the same developmental process. At A1 and A2, stronger teacher control is not a limitation on personalization; it is what makes personalization safe and comprehensible. At B1 and B2, greater learner freedom is productive only when learners have developed the linguistic and critical resources to use it. This helps reconcile two priorities that are sometimes presented as opposites: individualized support and teacher-led pedagogical coherence.

The same perspective also reframes the relationship between teacher and AI. AI can assume some routine functions—generating additional examples, varying practice, or offering first-pass feedback—while the teacher concentrates on interpretation, interaction, cultural meaning, motivation, and higher-order feedback. This is not a simple transfer of labor from teacher to machine. It is a redistribution that requires the teacher to design the boundaries of automation. The more fluent and persuasive AI becomes, the more important professional judgment becomes in deciding when fluency is pedagogically useful and when it conceals a problem.

The insistence on social reintegration is particularly important. Personalized digital practice is often imagined as an individual activity. Yet language competence is fundamentally relational. Learners need to experience misunderstanding, turn-taking, disagreement, repair, humor, hesitation, and the need to adapt to another person. AI may help learners prepare for these moments, but it should not remove them. The most promising model is therefore hybrid: individualized technological support feeds into shared human activity.

This orientation also makes the model more resilient to rapid technological change. Research will continue to evaluate new systems, and their capabilities will evolve. A durable pedagogical framework cannot depend on the assumption that a particular model will always make the same errors or have the same strengths. What can remain stable are the educational questions: What is the learner trying to learn? What should the learner do independently? What assistance is appropriate? How will the output be checked? Where will human interaction occur? What evidence will show that learning, instead of task completion, has taken place?

Because this is a conceptual chapter, it does not claim causal evidence that the proposed model itself produces gains in proficiency. Its value lies in organizing existing concerns and affordances into a coherent set of decisions for German preparatory education. Future empirical work should test stage-sensitive tasks across proficiency levels, examine how different feedback arrangements affect independent performance, and explore how learners' AI literacy develops over time. Longitudinal research is

particularly important because short-term improvements in a revised product may not translate into durable language development (Godwin-Jones, 2024; Pham & Renandya, 2026).

Conclusion

AI can undoubtedly make German preparatory education more flexible and responsive. Neither quality, however, is a pedagogical achievement in itself. What matters is whether the support fits the learner's present competence and still leaves room for effort, judgment, and interaction. Between A1 and B2, learners do more than accumulate vocabulary and grammar. They gradually become able to sustain discourse, read context, make pragmatic choices, evaluate feedback, and take greater responsibility for their own language use.

For beginners, the implication is straightforward: AI should remain closely tied to clearly defined learning goals. Controlled vocabulary practice, pronunciation rehearsal, grammar reinforcement, and short dialogues can provide useful extra practice when the language has been checked and the task remains manageable. At B1 and B2, the same technology can be used less as a source of answers and more as a resource for questioning: comparing formulations, testing a line of argument, examining register, revising a text, or preparing for extended interaction.

Across the progression from A1 to B2, the teacher is the person who keeps the larger pedagogical picture coherent. Teachers decide what matters, how difficult a task should be, when generated material is trustworthy enough to use, and when students need to leave the screen and talk to one another. They also make visible a lesson that is easy to forget when AI writes fluently: a convincing answer is not necessarily a correct or appropriate one. Learners, for their part, need opportunities to develop the confidence to question suggestions, reject them, and explain their own choices.

The aim, then, is not to give AI progressively more space simply because learners have reached a higher level. Its role should become more intellectually demanding. A useful progression moves from repetition toward reflection, from accepting corrections toward weighing alternatives, and from dependence on automated support toward informed agency. Used in this way, AI can widen the range of practice available in German preparatory education without displacing the teacher, the peer group, or the learner's own responsibility for learning.

Table 1. Stage-Oriented AI Personalization from A1 to B2

Stage	Primary learner need	Appropriate AI role	Teacher mediation	Learner responsibility
A1	Foundational vocabulary, basic patterns, confidence	Controlled repetition, pronunciation models, short drills and dialogues	**High: preselect language, constrain prompts, verify output**	Attempt short forms; compare with approved models
A2	Expansion of routine communication and grammar	Targeted practice, guided dialogue, simple feedback	**High to moderate: calibrate complexity and explain corrections**	Notice patterns; ask limited follow-up questions
B1	Coherent texts, interaction, strategy use	Guided revision, comprehension support, rehearsal	**Moderate: define purpose, monitor feedback, connect to peer work**	Produce first; evaluate suggestions; justify revisions
B2	Argumentation, register, mediation, pragmatic judgment	Comparative feedback, counterarguments, register analysis, complex simulation	**Strategic: challenge interpretations, support cultural/pragmatic evaluation**	Verify, reject or adapt output; defend independent choices

Declaration of Artificial Intelligence Usage

In preparing this work, I used various artificial intelligence (AI) applications to support proofreading, language and style editing, text refinement, preliminary research, brainstorming, and draft organization. All AI-assisted content was critically reviewed, verified, and revised by me. I take full responsibility for the originality, accuracy, academic integrity, and final content of this work.

References

Fuchs, B. (2026). Prompting as a metacommunicative competence in an AI-mediated foreign language classroom. Fremdsprachen und Hochschule, 102. https://doi.org/10.46586/fuh.v102.2025.12725

Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. arXiv. https://doi.org/10.48550/arxiv.2403.20216

Harsch, C., Müller-Karabil, A., & Buschmann-Göbels, A. (2025). Considering generative AI in language education in HE: The perspectives of learners, teachers and the language centre. Language Teaching Research Quarterly, 51, 184–202. https://doi.org/10.32038/ltrq.2025.51.01

Hartmann, D. (2024). Künstliche Intelligenz im DaF-Unterricht: Mediation mit KI-Anwendungen fördern. Revista Lengua y Cultura, 5(10), 50–55. https://doi.org/10.29057/lc.v5i10.12218

Häußler, A., Mersmann-Hoffmann, H., Richter, S., Allirand, L., Bukenberger, K., Engelhardt, M., & Poblete, J. S. (2026). Einsatz von KI-Systemen im Fremdsprachenunterricht an Hochschulen. Fremdsprachen und Hochschule, 102. https://doi.org/10.46586/fuh.v102.2025.12644

Joo, S. H. (2024). Generative AI as writing or speaking partners in L2 learning: Implications for learning-oriented assessments. Studies in Applied Linguistics and TESOL, 24(1). https://doi.org/10.52214/salt.v24i1.12865

Köylü, Y. (2025). Unlocking AI for language education: Mastering prompts, critical evaluation of AI responses, and implications for language teaching and learning. Journal of China Computer-Assisted Language Learning. https://doi.org/10.1515/jccall-2025-0017

Kul, A. N., & Canoğlu, H. D. (2025). Zum Einsatz von KI-gestützten Werkzeugen in Kombination mit Web 2.0 Anwendungen in der Didaktik der deutschen Sprache. Diyalog Interkulturelle Zeitschrift für Germanistik, 1127–1144. https://doi.org/10.37583/diyalog.1838317

Lembcke, B., Straube, A., & Jairoudi, K. A. (2025). KI im Sprachunterricht. Fremdsprachen und Hochschule, 102(1). https://doi.org/10.46586/fuh.v102.2025.12503

Malechová, M. (2025). Positionierung und Einflussnahme digitaler Medien und künstlicher Intelligenz im Fremdsprachenunterricht (pp. 145–155). https://doi.org/10.36689/uhk/978-80-7435-950-7.011

Meike, A. (2025). Kooperative Mensch-Maschine-Co-Kreationen? data_UMR. https://doi.org/10.17192/ziaf.2025.5.1.8799

Mohamed, M. S. P. (2024). Exploring ethical dimensions of AI-enhanced language education: A literature perspective. Technology in Language Teaching & Learning, 6(3), 1813. https://doi.org/10.29140/tltl.v6n3.1813

Pham, V. P. H., & Renandya, W. A. (2026). Navigating the AI turn: Framework for responsible AI integration in language education. International Journal of TESOL & Education, 6(2), 70–85. https://doi.org/10.54855/ijte.26626

qizi, Q. N. R. (2025). Challenges and limitations of using AI in language teaching. Zenodo. https://doi.org/10.5281/zenodo.18076770

Raveling, A. (2025). Search “A2, Liebesleben, 500 words”. Babylonia Journal of Language Education, 1, 22–27. https://doi.org/10.55393/babylonia.v1i.500

Renandya, W. A., & Floris, F. D. (2024). The changing landscape of ELT: Past lessons, future visions. PASAA, 69, 40–61. https://doi.org/10.58837/chula.pasaa.69.2

Strasser, T. (2020). Künstliche Intelligenz im Sprachunterricht: Ein Überblick. Revista Lengua y Cultura, 1(2), 1–6. https://doi.org/10.29057/lc.v1i2.5533

Xia, J., Ge, Y., Shen, Z., & Najar, M. R. (2024). The auxiliary role of artificial intelligence applications in mitigating the linguistic, psychological, and educational challenges of teaching and learning Chinese language by non-Chinese students. The International Review of Research in Open and Distributed Learning, 25(3), 116–133. https://doi.org/10.19173/irrodl.v25i3.7680

Yang, H., & Kyun, S. (2022). The current research trend of artificial intelligence in language learning: A systematic empirical literature review from an activity theory perspective. Australasian Journal of Educational Technology, 180–210. https://doi.org/10.14742/ajet.7492

Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students’ perceptions and preferences. Technology in Language Teaching & Learning, 6(3), 1482. https://doi.org/10.29140/tltl.v6n3.1813

AI-Driven Learning Analytics for French Language Materials Design across CEFR levels A1 to B1

Fayrouz Merabti [1]

Abstract
This chapter explores how AI-driven learning analytics connect learner data with the adaptation of French language materials across CEFR levels A1 to B1. Moving beyond fixed resources, it proposes a continuous, data-informed cycle where student interaction generates data, AI analyses the patterns, and teachers use these insights to evolve materials. The application scales with proficiency: identifying foundational difficulties for structured A1 materials, guiding communicative activities at A2, and supporting extended production, problem-solving, and autonomy at B1. Crucially, AI acts as a collaborative tool that supports rather than replaces teachers, who retain responsibility for ensuring pedagogical appropriateness, cultural relevance, and alignment with CEFR standards. Finally, addressing critical challenges such as data privacy, bias, and the limits of measurable indicators, the chapter argues that analytics should inform rather than dictate design. Ultimately, it advocates shifting from designing materials for learners to designing materials with evidence from learners, merging AI efficiency with teacher expertise to create truly adaptive resources.
Keywords: artificial intelligence; learning analytics; French language education; materials design; CEFR

Introduction

AI has progressed from a nascent technological opportunity to an increasingly prominent part of learning practice. AI systems are employed in language education for conversational practice, feedback, content production, language analysis, assessment, personalisation, and designing learning activities. A recent scoping review of generative artificial intelligence (GenAI) in language learning and language teaching has revealed a developing research stream as well as unanswered questions about effectiveness, ethics, specific language interventions, stakeholders, and teacher development (Law, 2024). More recent literature shows that although AI research in language learning has grown significantly since the advent of generative AI, it has been limited to certain languages, skills, learner groups and research designs (Wang et al., 2026). While these developments

[1] Lecturer, Istanbul University-Cerrahpaşa, School of Foreign Languages, Department of Foreign Languages, French Preparatory Program, ORCID: 0009-0009-7733-8601, *fayrouz.merabti@iuc.edu.tr*, Türkiye

seem to make AI increasingly relevant to the design of language materials, they do not, by themselves, make its role in education obvious.

The question raised in this chapter is, then, not whether AI can create language-learning materials. Modern AI programmes can create exercises, explanations, dialogues, prompts, texts, questions, feedback and alternate versions of learning activities. The research on AI in foreign language education has been a vast area which lists a variety of functions such as language practice, content generation, voice and audio applications, text generation, feedback, gamification, and course development (Kır & Tekin, 2025). A more significant question is whether these technological capacities can be linked systematically to evidence that can be found of what learners are doing and what they are learning. The pedagogical problem is not just about how to produce more, it's about what, for whom, why and based on which evidence.

One potential connection between learner activity and pedagogical decision-making is provided by learning analytics. Traditionally, the measurement, collection, analysis, and reporting of data related to learners and learning environments have been the focus of learning analytics (Hernández-de-Menéndez et al., 2022; Yılmaz & Çakır, 2021). A wide variety of analytical techniques and possible applications have also been identified such as visualization, semantic analysis, educational data mining, prediction, clustering, curriculum development, personalized learning, and analysis of learner behaviour (Nunn et al., 2016). The literature also points to methodological, ethical, interpretive, and privacy issues with it, however. So, it is not a rule that the more learner data, the more pedagogical knowledge.

This separation is especially crucial when paired with AI in the context of learning analytics. AI can accelerate the speed and scale at which learner data can be processed and can assist in the creation of potential instructional responses. The recent meta-analysis indicates that the benefits of AI-supported language-learning interventions can be positive for multiple language outcomes, but with varying degrees of impact depending on various contextual, instructional, and social-emotional moderators (Wu, 2024). Meanwhile, a 20-year review of the literature revealed significant gaps in the research, such as small sample sizes, brief interventions, a focus on English reading and writing skills, lack of diversity in contexts, and methodological opacity (Wang et al., 2026). It does not mean that there is no pedagogical value in AI. Instead, its value can only be determined within a specific context, design, implementation, and interpretation.

This chapter builds on this insight and proposes a specific argument: How pedagogically valuable the learning analytics provided by AI are depends not so much on the amount of automation the system allows, but on the quality of the relationship between learner evidence, AI-supported interpretation, teacher judgement and pedagogical action. When it comes to discussing educational AI, there is a fine line between four phases of the argument, three of which are frequently conflated: data, analytics, interpretation and pedagogical action. Learner activity generates data, patterns are found in the data, teacher interpretation of whether or not the patterns are pedagogically meaningful, and the materials are adapted based on a pedagogical decision.

The chapter therefore suggests an AI-based, CEFR-oriented materials adaptation cycle for teaching the French language. The learner interaction is placed at the start of the cycle, and the learner reengagement is placed at the end of the cycle of the model. AI provides computational analysis and generative ability between these two points, while the teacher is responsible for interpretation and pedagogical decision making. The model is used at all three levels of the CEFR (A1, A2 and B1) but cannot be viewed as a hierarchy of algorithmically determined learner profiles; rather, it is one of progressively different pedagogical contexts where different types of evidence have significance. The chapter is conceptual and does not report the results of an empirical study. The adaptation cycle outlined in this chapter is offered as a conceptual model that introduces the possibilities of using AI tools for learning analytics to inform the design and refinement of French language learning resources. It serves as a framework for future application and empirical study from a theoretical background.

Rethinking Learning Analytics in Language Education

From Learner Data to Pedagogical Evidence

Learning analytics is frequently associated with the collection and analysis of learner data. But it's not all about data collection. The term learning analytics is defined as the measurement, collection, analysis and reporting of data related to the learner and the context of learning, and the use of such data to better understand and improve learning processes (Hernández-de-Menéndez et al., 2022). Likewise, Yılmaz and Çakır (2021) highlight the potential of learning analytics in the comprehension of learners' behaviour, individualized learning, and decision making in education. The following definitions lay the foundation for the current argument: Learner data have educational significance when they are interpreted in relation to learning and teaching.

In language education, the difference between data and evidence is important. When a learner selects a wrong form of a word, such as a wrong grammatical form, it can be documented as an observable pattern, if this happens repeatedly. The pattern does not, however, necessarily contain information about the reason for the learner's production of the form. The cause of the error may be due to a failure in the grammatical knowledge, to the influence of the first language, to the difficulty of the task, to the lack of concentration on the task, to the unfamiliarity with certain words, to a lack of processing space, to a failure to understand the challenge, or to several factors. The pattern may be made visible using learning analytics, but this visibility does not mean that it is explained.

A conceptual distinction between learner data and pedagogical evidence is thus proposed in this chapter. Learner data are a representation of learner activity. Those representations are organized or analysed by Analytics. Pedagogical evidence is produced by the observed pattern and by making connections with a learning objective, instructional context, learner history, and task characteristics and plausible explanations for the observed behaviour. The difference is that it continues the existing purpose of learning analytics and doesn't suggest that analytics are sufficient to reach the pedagogical understanding. This is in line with the constraints found in the learning-analytics research. Nunn et al. (2016) recognise, for instance, the potential gains and the issues of tracking, collecting, evaluating, analysing, ethical, privacy and the relationship between learning analytics and learning sciences. The definition of value of the analytical output is not only computational sophistication, as suggested by these challenges. The question in the education is what is the output for this learner in this task at this point in the learning process?

The Problem of False Precision

One of the most challenging aspects of AI systems is the impression of precision that can accompany complex algorithms; the danger is that this precision might not always align with the accuracy of the interpretation. The dashboard can show a percentage, prediction, proficiency estimate, or a recurring error pattern. The numerical or algorithmically generated information can seem more objective than the qualitative observation by a teacher. But uncertainty persists with the numerical representation. This problem may be viewed as false precision, or the creation of highly specific appearing outputs from evidence that may not be complete, may be context dependent, or may be ambiguous. The idea is advanced here as a critical development of the methodological issues identified in learning-analytics research.

Because analytics are based upon what is measured and how it is measured, the interpretation that results are necessarily influenced by the design of the measurement system (Nunn et al., 2016; Hernández-de-Menéndez et al., 2022). This is a problem particularly with respect to language learning, since not all learning is equally quantifiable. Answers to automatically scored tasks may be evidence of performance of specific items but do not necessarily reflect interactional flexibility, pragmatic competence, strategic competence, willingness to communicate, mediation or the ability to transfer knowledge into new contexts.

The CEFR itself has linguistic, sociolinguistic and pragmatic aspects, and involves linguistic production, interaction, mediation and online interaction (Council of Europe, 2020). Thus, a materials-design system that takes into account only easily measurable variables may develop an incomplete model of L2 acquisition. Here, it is not the case that measurable data are unhelpful. On the other hand, they can be very beneficial. The argument is that measurement is a means of evidence and not an entire picture of learning. Combining quantitative patterns with teacher observation, learner work, task analysis, classroom interaction and the goal of the instructional programme is pedagogically stronger for the design of materials using AI.

Learning Analytics in Language Learning

Language learning patterns that are hard to identify can be uncovered by learning analytics. When used at scale, it can highlight errors that recur, patterns of responses, patterns of participation, completion of tasks, vocabulary challenges, or variations in task performance. The possibilities are consistent with the wider applications of learning analytics found in other studies (Nunn et al., 2016; Yılmaz & Çakır, 2021). AI may be able to broaden this ability by learning more complex sets of diverse language inputs and produce candidate interpretations or instructional resources. But what has been observed in the AI literature in general is that it is not a given. The results of the meta-analysis conducted by Wu (2024) indicated that the impact of AI interventions could differ based on the learning context, intervention length, educational level, learner attributes, language proficiency, interactivity, and feedback sources.

The finding has implications for materials design because it calls into question the notion that the use of a computer-based instrument is beneficial by itself, without regard to how it is pedagogically designed. Wang et al. (2026) also point out the methodological shortcomings in the existing literature on AI in language learning. Their review concluded that they focus on English, productive skills, university learners, relatively small

samples, and short interventions, and that they look for more varied contexts and better research designs. For French materials design it is a direct consequence. The evidence that comes mainly from English-medium university contexts must not be assumed to be applicable to the French learners in other educational contexts.

This chapter thus considers AI analytics as a form of "situated evidence" rather than a "universal" learner knowledge. AI can respond to the question, "What is the trend that can be seen from the data that is available?" The pedagogical process still needs to consider another question: What could this pattern imply, and what instructional response is warranted?

Artificial Intelligence and the Transformation of Language Materials Design

From Fixed Materials to Responsive Materials

Conventional language materials are normally developed beforehand and are not specifically tailored to the individual learners they will serve. Textbooks, worksheets, exercises and evaluation tasks are designed around expected curriculum goals and learner traits. This is an organization that gives stability, progression and coherence. It also establishes an inherent gap between the assumption that underlies materials and the need of a specific group. AI and learning analytics have the potential to bridge some of that gap. Learner activity can be analysed in the course of instruction, and the materials can be adapted based on it. This in principle renders materials more responsive. However, responsiveness is not synonymous with personalisation for each child.

When each learner is presented with a different sequence of tasks, the curriculum can be split, and the common learning space can be reduced in effectiveness. According to Law's (2024) review, there are several areas in which GenAI has been used in language education, and the review also pinpointed the need for further empirical studies, ethical analysis, specific interventions, and professional development. Similarly, Kır and Tekin (2025) illustrate the host of applications of AI in foreign language education. These findings contribute to the technologic feasibility of responsive materials, but not the pedagogic desirability of responsiveness at all times. Conceptually it is to reconceptualize responsiveness as principled adaptation. Materials should be modified when there is enough evidence that they will be better for supporting an identified learning objective. It's not about personalization for the sake of personalization. It is appropriate integration into an integrated program.

Generative AI From Content Production to Pedagogical Design

The advent of GenAI has opened up new possibilities to enhance the pace at which teachers are able to create and shape language-learning material. A teacher may ask for alternative materials to accompany the text, more examples, vocabulary practice, comprehension questions, dialogues, role play prompts, or different versions of a task. According to research reviews, content creation is one of the vital areas in which GenAI can be implemented in language education (Law, 2024; Kır & Tekin, 2025). However, content produced at speed is not always top-quality content design. An activity that is generated can be grammatically correct but pedagogically not. It can include words at a higher or lower level than the task required, miss the sense of a task's communication, misrepresent the actual purpose of communication or include examples that are culturally inappropriate.

In the same vein, Son and Philpott (2026) therefore position AI-powered language teaching in terms of: not only the tools and benefits, but also the limitations and skills of the language teacher. So, the difference lies between the production of content and pedagogical design. Pedagogical design involves choices about goals, sequencing, cognitive complexity, language appropriateness, interaction, assessment, and the learner context in designing the candidate materials; AI can generate materials. The chapter's model, therefore, regards AI-generated content as tentative teaching materials to be evaluated by professionals.

Why Automation Cannot Replace Pedagogical Judgement

When AI is used in the design of teaching materials, the role of the teacher is more vital, not less. The people who design, implement and mediate AI in education play a crucial role in shaping its transformative potential, and teachers' professional capacities and agency are highlighted by Cukurova (2025). The report also introduces an explicit focus on professional learning related to using AI to augment and not supplant pedagogical knowledge and ethical decisions. At the same time, Son and Philpott (2026) centre the roles and competencies of teachers as a key element of language teaching through artificial intelligence and highlight the need for teacher-student interaction to be critical, effective, efficient, and ethical.

Continuing professional development is also crucial for informed decision-making and effective integration of GenAI in language education, as noted by Law (2024). Drawing on this, the present chapter introduces the concept of the teacher as an interpreter of evidence generated by AI. This

distinction matters. When teachers merely follow AI recommendations, their roles shift towards some part of the decision-making process being taken off their hands, but the outputs of AI systems may be opaque or contextually incomplete. When teachers interpret the outputs of AI, however, the technology itself is a source of evidence in professional decision making.

The CEFR as a Framework for Interpretation, Not Automation

CEFR as a Framework for AI-Driven Materials Design

The Common European Framework of Reference for Languages (CEFR) is a broad framework to be used for curriculum development, teaching, language materials and assessment. According to the Council of Europe (2020), language competence is defined in an action-oriented approach that includes linguistic competence, sociolinguistic competence and pragmatic competence, and actions like producing, interacting, mediating and online interacting. The CEFR therefore goes beyond being a continuum of language knowledge and provides a more comprehensive framework of reference for the design and development of language materials. This is especially relevant in the context of AI-based materials design; AI-based adaptation needs to be aligned with meaningful language use goals, and not just to the number of errors or linguistic forms. For instance, one learner with a frequently incorrect grammatical form might need more form practice while another learner at the same proficiency level might need more help with interaction, discourse structure, mediation, or pragmatics.

Proficiency levels in the CEFR, however, should not be used as a complete psychological or pedagogical profile of an individual learner. Two learners with a general level of B1 can have very different profiles in terms of their abilities and difficulties: one learner could have relatively good grammatical competence but poor fluency, while another learner could have relatively high fluency but relatively low grammatical competence, or could have good fluency but difficulties with accuracy or with organizing the discourse. The multidimensionality aligns with the CEFR (Council of Europe, 2020) and findings that the efficacy of AI-assisted learning may differ depending on conditions of the learner and the learning situation (Wu, 2024). Therefore, pedagogical constraints and contexts should be used to limit and shape AI-driven materials adaptation, rather than relying on an algorithm to automatically identify learner needs.

The CEFR outlines general guidance on language activities and language competencies associated with the different levels of proficiency, but

learning evidence can inform teachers to identify the specific areas that need further support and teacher interpretation is always crucial to inform teachers about the pedagogical needs.

Reconsidering A1–B1 Progression Through Learner Evidence

The A1, A2 and B1 levels should not be singled out as separate areas of learning grammar skills and autonomy as this may disguise the more complex concept of the CEFR. Rather, the focus might vary in different levels of competence, and several aspects of communicative competence might be in play throughout the process. A1 can have a greater focus on relating the high-frequency forms, meanings, vocabulary, comprehension, pronunciation and basic communicative functions of materials. A2: Learners need to make more frequent connections between language resources when communicating in their routine exchanges and to create more coherent language when communicating in longer stretches. B1: Materials can facilitate longer production, interaction, mediation, discourse organization, lexical range, and problem-solving. They are not to be interpreted as distinctive attributes of specific levels, but rather as evolving aspects of communicative competence (Council of Europe, 2020). In this trajectory, AI-powered analytics can be used to detect factors that might need further focus on the learner's performance. The system must not, however, presume that all A1 learners need to be supported in the same way, or all B1 learners in another way. Proficiency level in the broader pedagogical horizon and learner evidence in the instructional route.

From Learner Interaction to AI-Supported Evidence Interpretation

The proposed adaptation cycle starts with the interaction of learners with texts, where the learners interact with the texts, do activities, participate in conversations, write text, do vocabulary activities, do assessments, do online activities, and interact with texts in other ways, all observable and measurable. The nature of the activity is critical, because the activity that creates learner data influences the nature of the data. A learner may struggle with a grammatical construct in a non-contextual task but use the same linguistic construct differently in a meaningful communication. Thus, the performance of the learners is not independent of the context and the activity where it was obtained. The next step is to convert the raw activity data into evidence of learning that is both meaningful and educational. Learning analytics can offer the organisation of answers, detection of recurring errors, comparison of performance between tasks, and analysis of variations over time, as in the past (Nunn et al., 2016; Hernández-de-Menéndez et al., 2022).

The model does not assume that all miscues are "evidence" that needs to be addressed right away but instead focuses on building and analysing evidence over several tasks and contexts. The capabilities of AI can help in this process, as they can identify patterns within large sets of learner evidence, locate recurring linguistic patterns, arrange learner generated texts, and propose potential explanations or instructional suggestions. The growing body of literature on language learning with the aid of AI and personalization shows the potential of these applications (Wu, 2024; Wang et al., 2026). Nevertheless, pattern recognition should not be equated with diagnosis.

There could be a pattern of repeatedly using the same form incorrectly, but this does not indicate why the problem is arising, or what pedagogical intervention might be helpful. The conceptual cycle can then be expressed as: CEFR-informed objectives → initial materials → initial learner interaction → initial learning evidence → initial AI-supported analysis → initial teacher interpretation → initial pedagogical decision → adapted materials → engagement of the learner again, resulting in renewed learning evidence. This series sees AI not as an independent decision maker, but as a support tool for evidence processing and pattern recognition.

Teacher-Mediated Materials Adaptation as a Cyclical Process

The interpretation done by the teacher will play a pivotal role in bridging the gap between the AI-generated analysis and the pedagogical action. The teacher reflects on patterns identified when considering the learner, the task in which the evidence was produced, the curriculum, the CEFR objective, previous learning and the overall classroom context. This position is aligned with the notion of teacher agency in AI-supported education (Cukurova, 2025) as well as the essential skills needed to effectively and critically engage in AI-powered language learning (Son & Philpott, 2026). The teacher can accept the interpretation, refuse it, demand more evidence, ask for more information about the learner or choose another pedagogical response. In this model, a failure to agree with AI is not a failure of the system, but rather a professional teacher choice. Interpretation, then the teacher decides on any changes to the materials.

Adaptation can be through more practice, task difficulty, alternative text input, more scaffolding, more communication opportunities, alternative examples, peer interaction or changing the type of task. It doesn't need to be a new creation and can involve ordering, combining, adding, shortening, contextualizing, and/or augmenting existing materials with other appropriate materials and/or resources. While AI can aid in these adjustments, the pedagogical principles continue to rely on teacher judgment and evidence from learners. The adapted materials are then presented to

learners, and a new round of interaction and evidence follows. Materials adaptation is therefore not linear but rather cyclical as the effectiveness of a pedagogical decision is investigated by the learners' activity following the decision, not just its plausibility when it was made by an AI-based tool. In case of an unanticipated change, the teacher can go back to the interpretation, to the pedagogical decision, and change the materials once more.

The cycle then includes an ongoing process of pedagogical inquiry that is embedded in materials design with AI, running from CEFR-informed aims, learner evidence, AI-supported analysis, teacher interpretation, and back to materials adaptation.

Applying the Model Across A1, A2, and B1

Foundations, Scaffolding, and Controlled Expansion

Students are building basic language resources and beginning to use basic communicative functions. At this stage materials may then need a great deal of scaffolding and controlled progression. The CEFR descriptors offer a benchmark of the types of language tasks that are linked to this level (Council of Europe, 2020). In other words, it requires materials to build sufficiently accessible connections between form, meaning, vocabulary, pronunciation, comprehension and basic communicative functions before learners are expected to proceed independently. In other words, progression at A1 level is not just the addition of more complex grammatical structures, but also the ability to identify and make use of basic resources in relevant yet manageable contexts.

The advantages of using AI-driven analytics to spot recurring issues with basic communicative functions, comprehension activities, high-frequency vocabulary and foundational forms can be especially beneficial. This potential should, however, be approached with a degree of caution as the reoccurrence of an error does not necessarily mean that the learner needs further practice with the same item. A recurring problem could be due to a lack of conceptual understanding, lack of contextual support, unfamiliar task format, and/or a disparity between what the learner has and what the activity requires (Council of Europe, 2020). The repeated error should not lead to repetitive drilling, therefore. When a learner has multiple errors with an item, an alternative approach could be to alter the explanation, representation, task context or level of scaffolding. The evidence produced by AI should then be interpreted as a signal to need more pedagogical interpretation, rather than as a command to do more practice.

From Controlled Performance to Functional Communication

Learners start to use language in routine interactions and connected activities at A2. This raises the question of looking at correctness along with functional communication. The action-oriented nature of the CEFR does not lend itself to the idea of a list of grammatical forms to be accumulated at A2 level (Council of Europe, 2020). AI can aid in discovering trends in the writing, answers, vocabulary, or interactional activities of learners. Wu (2024) presents evidence that intervention with the assistance of AI has impacts on several language outcomes and that impacts differ across instructional and contextual situations. This further emphasizes the importance of tying analytics to task design instead of making AI effectiveness a standalone concept.

For A2, this may mean providing more opportunities for meaningful interaction, more lexical resources, more development of connected discourse, or changes to the support of the tasks. Teacher decides which response is appropriate.

Extended Production, Interaction, Mediation, and Autonomy

At B1, the pedagogical picture becomes more and more complicated as the learners are expected to work independently in longer language tasks. The CEFR defines areas in which the language may be described, such as interaction, mediation and online interaction (Council of Europe, 2020). At this stage, the use of AI-based analysis could be helpful in analysing patterns over longer texts, multiple writing tasks, oral production, interaction, and learner reflection. The longer and more complex the performance, the greater the interpretive problem, however. A weakness in an extended text could be attributable to limitations in vocabulary, limitations in discourse organization, task demands, unfamiliar content or several interacting factors. Therefore, the emphasis should be given to triangulation in B1 materials adaptation. Teacher observations, learner products, task characteristics, and learner self-reflections should be taken into account when using AI-generated patterns.

The Teacher at the Centre of Pedagogical Decision-Making

From AI Output to Pedagogical Interpretation

The use of AI in language learning should not be seen as a shift from teacher to machine in teaching. Instead, it alters the nature of the teacher's job, as it brings new types of evidence, automated analysis, and content generation that need professional interpretation. The core of the learning analytics field is the measurement, analysis and reporting of information

on learners and their contexts to understand and enhance learning (Hernández-de-Menéndez et al., 2022). In the field of foreign language education, the applications of AI can be used in many fields such as foreign language training, content writing, feedback, speech applications, and personalized education (Kır & Tekin, 2025). But these functions are not necessarily pedagogically suitable just because they are available. The real question is not what AI can discover in patterns or create in materials, but what does it represent, educationally speaking, and what should be done about it.

This difference is especially significant when it comes to designing materials with artificial intelligence. A valid pedagogical decision does not necessarily stem from an AI system identifying repetitive errors, creating a learner profile, suggesting an activity or producing an alternative text. The development of learning analytics has long acknowledged the limitations of being able to use available data to inform an educational interpretation, and systematic reviews of the use of AI in learning analytics in education have identified substantial potential, yet with some methodological and implementation hurdles (Hernández-de-Menéndez et al., 2022; Salas-Pilco et al., 2022). The findings from more recent language education studies also indicate that the effectiveness of the learning outcomes through AI is dependent on the context and the teaching and learning environment.

The results of recent research in language education also reveal that the learning outcomes that can be achieved through AI are not solely determined by the technology, but also by the context and the teaching and learning environment. The teacher is then in an important interpretive position in the link between the computations and the teaching action. This chapter thus frames the teacher as an interpreter, evaluator, and pedagogical decision-maker in addition to being an AI user.

The difference between an AI recommendation and pedagogical decision is critical as there is a difference between what can be recommended and what can be prescribed. When the AI system repeatedly catches a grammatical structure that students struggle with, it could be a reason to investigate further, but it does not always mean that there is a need for more grammar exercises. The problem can relate to both task design and inadequate contextualization; vocabulary demands and comprehension problems and/or to differences between written and spoken performance. On the other hand, no errors do not imply successful communicative competence. The challenge is not that AI can't find patterns but that the patterns become pedagogically relevant when they are linked to learning goals, task parameters, student characteristics, and curricular expectations. The

findings align with the current body of research, as recent reviews have emphasized the need for further empirical research and methodological limitations of the studies (Law, 2024; Wang et al., 2026).

Teacher Expertise as a Critical Safeguard

The role of teacher expertise is especially critical because language learning has aspects that may not be sufficiently captured by a few numerical indicators. Greater visibility of certain aspects of learner behaviour is possible through the expansion of learning analytics, but it is not a complete representation of learning. Nunn et al. (2016) highlight issues regarding event collection, analysis and interpretation of learning analytics, and Hernández-de-Menéndez et al. (2022) point to the difficulty of the process of converting learner data into information to support learning enhancement. The pedagogical risk is not just that the data might be wrong – it's the risk of measurable behaviour being seen as meaningful learning.

This is especially important for the development of language materials. Kır and Tekin (2025) list a wide variety of applications of AI in foreign language instruction, such as language practice, language question answering, speech-related applications, content generation and course development. These examples illustrate how AI can be applied across various facets of language learning, underscoring the critical need for informed pedagogy in choosing the right tools and techniques for the job. Technology capability is not a sign of educational appropriateness. In a similar manner, Son and Philpott (2026) propose a definition of AI literacy that goes beyond the mere use of the tool to critical and ethical reflection on the use of AI in language learning. This entails that instructors must assess the accuracy of the language used in the material provided by the AI, its relevance and suitability for the context, and its alignment with the learning objectives.

Teacher agency should therefore be seen as a core part of designing materials with AI support and not as a measure taken as an afterthought once the technology has been introduced. Cukurova (2025) points out that AI cannot replace teachers' professional skills and moral conviction. This is especially vital with the use of AI in the interpretation of learner evidence. A teacher can accept, adjust, or request more evidence, or reject the AI interpretation. This professional judgment is not an expression of technology resistance; it's a responsible action in the use of technology. The function of AI then becomes more about providing the teacher with access to potentially useful evidence, not taking away the pedagogical authority from the algorithm.

However, the same rationale applies to the interpretation of materials created by AI. In a recent scoping review on the use of GenAI in language teaching and learning, Law (2024) points out that the potential applications are great, but there is a need for more empirical studies and responsible use. In the same way, Wang et al. (2026) show that the research on the use of AI in the language learning field has seen significant growth with uneven research distributions by language, context, methodology, and outcome. Such restrictions emphasize the need for professional assessment. A produced French activity can be grammatically correct but not suitable for the purpose of communication, the culture, the level of the learners, or the cognitive aspects. Thus, generation should not be confused with materials design. AI can speed up the production process, but teaching quality will still need to be assessed by experts.

Critical Challenges and Ethical Tensions in AI-Driven Materials Design

Data Privacy, Governance, and Institutional Responsibility

The relation of learning analytics in materials development invariably deals with the collection and interpretation of learners' data, which calls for privacy and ethical governance as key issues. Previous learning analytics studies highlight privacy, data management and ethics as one of the key challenges in the field (Nunn et al., 2016). Ethics in the study of AI and LAs in-teacher education is another issue noted by Salas-Pilco et al. (2022). As AI systems handle more information and data types from learners, these concerns grow in importance. It is not only a question of whether data are technically protected, but also about whether the data collected are needed and proportional to the pedagogical goal, for how long they will be kept, who will be able to access them, and how decisions made based on the collected data could impact learners.

AI governance is a topic that has been studied recently, and it is important to think about these questions at the institutional level. Chiu (2026) points out the discrepancy in the speed of development and implementation of AI in the education sector and the development of governance frameworks, highlighting algorithmic bias, data sovereignty, transparency and accountability. These worries have immediate ramifications for AI-based materials development, as no single educator can be held accountable. Additionally, when using AI technologies for learning analytics, institutions must have processes to manage data, maintain transparency, provide professional learning opportunities, hold accountability, and challenge the

results of automation. Designing materials for responsible AI is thus a classroom and institutional responsibility.

Personalisation, Fragmentation, and Learner Agency

Personalisation is a one of the most talked about advantage of AI in education, and empirical studies have shown that certain AI interventions, such as adaptive and personalised learning tools, can positively impact language learning within specific contexts (Wu, 2024). The personalised learning application of AI in foreign language learning is also mentioned by Kır and Tekin (2025). However, personalisation shouldn't be seen as a good for education. When materials are constantly adapted to very specific student performance measures, learners may find their learning experience increasingly broken up into smaller pieces with less opportunities to interact in the classroom. The goal should thus be differentiated support instead of endless personalization: pupils can be offered various ways of support without having to achieve different objectives in terms of the curriculum and communication.

Agency for learners needs to be taken as well. Law (2024) notes the opportunities and challenges of GenAI in language teaching and learning, in this regard, Son and Philpott (2026) as well highlight the critical and ethical aspects of AI literacy in language education. It is, therefore, inappropriate for learners to solely be a receiver of AI-produced explanations, corrections, translations, or materials. Alternatively, materials can ask learners to question, confirm, compare, revise, and review the AI-generated content. These practices make the use of AI a part of language learning itself and foster learners' critical thinking about the reliability and suitability of AI-generated language.

Implications for French Language Materials Development

From Automated Personalisation to Evidence-Informed Adaptation

Drawing on the above, it is clear that AI-based materials development should be recognised as an evidence-based adaptation process rather than personalisation by the computer. Learning analytics can enhance the visibility of learner activity and can aid the detection of recurring patterns (Hernández-de-Menéndez et al., 2022; Yılmaz & Çakır, 2021). AI can then help to structure evidence, create potential explanations and generate candidate materials (Kır & Tekin, 2025; Law, 2024). But teachers need to assess the meaning of the evidence and whether the material proposed will meet the pedagogical need. This is in line with the literature reporting that

the learning outcomes of language learning with AI are different under different instructional and contextual conditions (Wu, 2024) and that the current evidence base is still fragmented in research contexts and methodologies (Wang et al., 2026).

For French language programmes, it is important that development of materials can be more responsive, though not fully automated. Learner evidence can suggest the need for: more vocabulary practice, more grammatical recycling, more support for pronunciation, more work on discourse, more interactional activities, more mediation work, or more communicative challenge. Learner evidence can help identify the level of support the learner needs to have additional communicative competences and activities, and the CEFR can serve as a reference for the determination of the wider communicative competences and activities relevant to the learner's level (Council of Europe, 2020). AI can help inform the identification and organisation of such evidence, yet it is still the teacher's responsibility to decide on the pedagogical response.

Materials development and evidence design should then be thought of as related. Tasks must produce evidence of learning that is meaningful if teachers will use learning analytics for informing materials adaptation. The evidence that can be provided for language use can vary by task type and should be understood as related to the design of the intended learning outcome. Materials, then, are not just a way to deliver content but are also means to create meaningful evidence of language development.

Implications for Teacher Education

This approach changes what teachers need to know about AI. For teachers, it's not their job to become AI programmers but it has become important that they understand the learner information being fed to AI systems, what they are outputs are, the uncertainty that still exists, and how their outputs relate to pedagogical objectives. Cukurova (2025) highlights teacher professional capacity and agency in response to AI and Son and Philpott (2026) view critical, creative, effective, efficient, and ethical use of AI in language teaching as essential elements. Therefore, teacher education must shift from training teachers on how to use tools to pedagogical AI literacy. Pedagogical AI literacy can encompass assessing AI output for adherence to the CEFR, checking for misinformation, or understanding whether the activities produced match authentic learners' communication purposes, comparing the analysis produced by the AI with real student performance, or designing learning activities that prompt learners to critically analyse the AI output. This way, the notion of AI competency is embedded in language teaching, not as an isolated language technology.

Implications for Future Research

The proposed framework also highlights some key areas for future research. Research should focus on French language learning and not just take for granted the results of English dominant research. Wang et al. (2026) point out key gaps in the existing literature regarding language distribution and language context and research method. Second, longitudinal studies are required as materials adaption is developmental and iterative, with changes in performance over short periods of time being insufficient to determine whether repeated AI-supported adaptation leads to long-term learning. Third, studies could explore teacher decision making, including teacher interpretations, adaptations, acceptance or rejection of AI-generated learner analyses and materials.

One of the most fruitful methods of the research is studying disagreement between AI and teachers. Disagreements should not be seen as evidence of technological problems or fault with the teacher. Rather, it could highlight the gap between computational pattern recognition and pedagogic interpretation. Exploring these crossovers could contribute to pinpointing areas where the French language performance can be accurately assessed with AI and areas that need more context or expert interpretation. This study also suggests that, in future, adaptive materials should not be assessed solely based on the quantity of AI-generated products, but also on their quality and pedagogical impact (Law, 2024; Kır & Tekin, 2025).

Conclusion

Using AI to analyse learning patterns can help make learner patterns more visible and offer evidence to guide instructional adaptations that can support the design of French language materials. But this potential cannot be realized unless the evidence is interpreted in a certain way. The CEFR serves as an important tool for establishing learning expectations at levels A1–B1 but is incapable of capturing individual learner development complexity. Likewise, AI-derived patterns should be interpreted as evidence that must be considered but cannot be assumed to be a true reflection of what the learner knows or needs. This is the central idea of this chapter effective adaptation emerges not only from the automation of the process but also from the interaction of learner evidence, the objectives set in the framework of the CEFR, and the teacher's competence and the design of the materials. This role aligns with studies demonstrating the capabilities and promise of AI and learning analytics in personalised learning and the ongoing issues of interpretation, context, ethics, and validity (Nunn et al., 2016; Law, 2024; Hernández-de-Menéndez et al., 2022; Wang et al., 2026).

This approach also emphasizes that teacher agency is integral to the development of materials with the help of AI. The teacher is still responsible for the meaningfulness of the identified pattern in the pedagogical context and what is the adequate response. Materials should thus be differentiated on a principled basis and not at any cost, taking account of the relevant differences of learners while keeping the shared aims of learning. Simultaneously, the implementation of this approach demands privacy protection, transparency, bias, and limitations of automated analysis (Britchenko, 2025; Chiu, 2026). AI literacy in language learning, therefore, goes beyond mere knowledge of how to use technological devices; it involves questioning the outputs generated by AI and using professional judgment to blend them with existing knowledge and skills (Cukurova, 2025; Son & Philpott, 2026).

Finally, the benefits of AI in the design of French language materials are not to supersede pedagogical judgement, but rather to augment the evidence that can inform it. Pedagogically interpreted adaptation, therefore, is proposed as a strategy that moves away from automated personalisation towards supporting teachers in recognizing and responding to learner needs while maintaining the agency of humans and learners. The future research is needed to study this approach at different proficiency levels, at different learners' groups, and in different situations of learning, while focusing on longitudinal evidence and methodological transparency (Wang et al., 2026). Within this paradigm, AI is not the one deciding on materials but is a means of enabling more thoughtful, critical, and responsive teaching and learning.

Declaration of Generative AI Use

ChatGPT (OpenAI) was used in the preparation of this chapter for language editing, organization, and revision. The author reviewed and approved the final content and takes full responsibility for the chapter.

References

Britchenko, I. (2025). Ethical AI in education: Principles, governance, and responsible implementation. *Pedagogy and Education Management Review*, (4(2)), 17–30. https://doi.org/10.36690/2733-2039-2025-4-17-30

Chiu, T. K. F. (2026). AI governance and ethics in education: Research gaps, future directions, and methodological approaches. *Interactive Learning Environments*, *34*(6), 3627–3630. https://doi.org/10.1080/10494820.2026.2710463

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of

Europe Publishing https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C. A., & Ramírez Mendoza, R. A. (2022). Learning analytics: State of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *16*(3), 1209–1230. https://doi.org/10.1007/s12008-022-00930-0

Kır, D. B., & Tekin, E. (2025). Key features of AI applications in foreign language teaching. *Bartın University Journal of Faculty of Education*, *14*(4), 1143–1162. https://doi.org/10.14686/buefad.1681050

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*, 100174–100174. https://doi.org/10.1016/j.caeo.2024.100174

Cukurova, M. (2025, September 4). *Promoting and protecting teacher agency in the age of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization (UNESCO). ResearchGate. https://www.researchgate.net/publication/395261486_Promoting_and_protecting_teacher_agency_in_the_age_of_artificial_intelligence

Nunn, S., Avella, J. T., Kanai, T., & Kebritchi, M. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2). https://doi.org/10.24059/olj.v20i2.790

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial Intelligence and Learning Analytics in teacher education: A systematic review. *Education Sciences*, *12*(8), 569. https://doi.org/10.3390/educsci12080569

Son, J.-B., & Philpott, A. (2026). AI-Powered language teaching. *Cambridge University Press 2026*. https://doi.org/10.1017/9781009650809

Wang, Y., Zhang, Z. J., & Zhou, H. (2026). Artificial intelligence in language learning: A twenty-year scoping review of applications, research methods, and outcomes. *Research Synthesis in Applied Linguistics*, 1–38. https://doi.org/10.1080/29984475.2026.2647961

Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126*, 103498. https://doi.org/10.1016/j.system.2024.103498

Yılmaz, F., & Çakır, H. (2021). Learning analytics and potential usage areas in education. *Journal of Learning and Teaching in Digital Age*, *6*(2), 81–89. https://files.eric.ed.gov/fulltext/EJ1308353.pdf

Artificial Intelligence-Based Dynamic Persona in Language Education: An Examination of Sociocultural Context and Pragmatic Competence

Burcu Ün [1] & Gizem Arslan [2]

Abstract
In foreign language acquisition, traditional teaching materials often present interactive dialogues through standardized and idealized prototypes that largely exclude social and emotional variables. This limitation may restrict learners' exposure to the sociocultural dimensions and linguistic variations of everyday communication. This study examines the potential contribution of the "Dynamic Persona," generated by Large Language Models (LLMs), to learners' pragmatic competence and linguistic adaptation. The research adopts Contextual Scenario Analysis as its methodology. Standard language-teaching contexts, such as restaurants, family, health, and work, are redesigned by assigning AI-generated characters specific traits, including grumpiness, introversion, and authoritarian attitudes. Through these variable character simulations, learners are encouraged to respond not only with grammatical accuracy but also by adapting forms of address, idiomatic expressions, and communicative strategies to the interlocutor's social status, personality, and emotional state. The study demonstrates the pedagogical potential of exposing learners to diverse personas and socially situated interactions, particularly for developing sociolinguistic awareness and communicative flexibility. Accordingly, artificial intelligence is conceptualized not merely as a static response system but as a flexible "sociocultural laboratory" in which learners and teachers can explore diverse communicative scenarios.
Keywords: Dynamic Persona; Large Language Models (LLMs); Foreign Language Education; Pragmatic Competence; Sociolinguistic Awareness

Introduction

The transformations that have taken place in the field of foreign language teaching in recent years have highlighted the need to approach language not merely as a set of grammatical rules, but also in terms of its social, cultural, and pragmatic dimensions. In particular, the Common European Framework of Reference for Languages (CEFR) defines the learner not

[1] Lecturer, National Defence University, School of Foreign Languages, Department of Western Languages, Italian Language Program, ORCID: 0009-0009-0397-8075, *burcu.un@msu.edu.tr*, Türkiye

[2] Lecturer, National Defence University, School of Foreign Languages, Department of Western Languages, Italian Language Program, ORCID: 0009-0006-9743-8536, *gizem.arslan@msu.edu.tr*, Türkiye

as a passive recipient of information, but as a "social actor" who performs tasks and actions in various social contexts. In line with this, the action-oriented approach that has been adopted does not limit language learning to the acquisition of linguistic structures alone; it also considers the individual's understanding of the target language community's cultural values, social norms, and communication practices to be an integral part of the learning process.

According to this approach, knowing a language does not simply mean learning vocabulary and grammatical rules. Learners are also expected to know in what context, with whom, and for what purpose to use the language, as well as to develop appropriate communication strategies in different cultural settings. For this reason, in today's foreign language education, communicative competence, pragmatic proficiency, and intercultural awareness have become goals just as important as linguistic accuracy.

However, when examining textbooks and traditional materials commonly used in foreign language instruction, it is evident that the dialogues presented to learners are mostly based on standardized and predictable communication situations. In real life, however, individuals interact with people who have different personality traits, different social roles, and different communication expectations. For this reason, learners must be prepared not only for linguistic structures but also for variable social situations and different forms of communication.

Generative AI systems, which have become widespread in the field of education in recent years, offer new possibilities in this regard. In particular, applications based on large language models can create virtual conversation partners for learners with different social identities, personality traits, and communication styles. Thus, the same communicative scenario can be reproduced with different contextual variables, allowing learners to experience situations that more closely resemble the unpredictable communication conditions of real life.

The aim of this study is to examine the potential of generative AI-supported dynamic persona applications in foreign language teaching and to demonstrate how these applications can contribute to the development of learners' communicative, pragmatic, and sociolinguistic competencies. To this aim, the study establishes a theoretical framework based on Hymes's approach—which centers on the concept of communicative competence—and Wittgenstein's understanding of language games, and evaluates dynamic persona applications within the context of foreign language teaching.

Research Model and Design

This study was designed as a qualitative research project aimed at identifying the structural and pragmatic differences between the static structure of traditional materials and the dynamic linguistic inputs generated by Gemini, a Large Language Model, in the foreign language acquisition process. The study adopted a comparative case study design, which allows for an in-depth examination and comparison of cases within their own contexts. Accordingly, the standard dialogue structures found in traditional teaching materials represent the first case, while versions of the same thematic contexts reimagined by artificial intelligence with specific psychological and sociocultural variables assigned by the AI represent the second case; both cases are analyzed within a linguistic framework.

Dataset and Corpus Creation Process

The study's dataset consists of four core method books prepared in accordance with the Common European Framework of Reference for Languages (CEFR) criteria and most widely used in Italian language instruction at the international level; these books are Via del Corso (Edilingua), Nuovo Espresso and Dieci (Alma Edizioni), and Spazio (Loescher). In accordance with the purposeful sampling method, units that most clearly reflect the sociocultural and pragmatic stratification across language levels were included in the corpus. In this context, restaurant interactions from the A1 level, health and doctor-patient relationships from the A2 level, and workplace dialogues from the B1 level were scanned and digitized, forming the static dataset that constitutes the study's control group.

Data Collection Tools and Dynamic Persona Construction

The dynamic dialogs that constitute the study's experimental dataset were obtained through prompt engineering processes conducted using Gemini. In the data collection process, drawing on Dell Hymes' theory of communicative competence and Wittgenstein's philosophy of language games as an epistemological foundation, original system prompts were designed to disrupt standard scenarios at every level of language and add socio-psychological depth to the text. In this vein, three distinct AI personas were constructed with the aim of moving beyond the linear, seamless, and idealized -sterile prototypes- devoid of social variables, as presented in traditional method books (Via del Corso, Nuovo Espresso, Spazio/Dieci). For the A1-level restaurant context, the model has been transformed into the persona of a hurried and impatient waiter working in a noisy Roman restaurant; thereby generating typical local reactions -in the form of dynamic variations- to tourists' morphosyntactically correct but culturally inappropriate inputs (such as requesting ketchup on pizza or ordering a

cappuccino at the wrong time), ensuring that students confront pragmatic failure firsthand. Throughout the process, in the A2-level healthcare context, the model was designed as an authoritarian pharmacist/physician character who maintains a hierarchical distance, creates legal/bureaucratic obstacles by refusing requests for over-the-counter antibiotics, and mandates the use of formal forms of address; In the B1-level job interview context, the model is assigned the persona of a skeptical interviewer who questions chronological gaps in the candidate's resume and uses vivid linguistic patterns and idiomatic expressions such as "chi si ferma è perduto." In short, this model aims to offer a dynamic "sociocultural laboratory" that transcends the linear and sterile boundaries of traditional textbooks, preparing students for the unpredictable sociopragmatic crises and lively linguistic interactions of real life.

Limitations of the Study

This study is limited, in terms of material, to nine specific units at the A1, A2, and B1 levels from the Via del Corso, Nuovo Espresso/Dieci, and Spazio textbooks used in Italian language instruction. From a technological perspective, the study is limited by the current algorithmic capacity, linguistic proficiency, and flexibility of the language model's responses to prompts during the data collection period. Finally, since this research does not include empirical testing of these generated dynamic dialogues on actual groups of learners or in classrooms, it is confined entirely to text analysis, material evaluation, and the proposal of a theoretical pedagogical model. The primary reason for limiting the study to the A1, A2, and B1 levels is to ensure that the dataset can be analyzed as a coherent whole and to keep the scope of the study under control. These levels were selected as the focus of the study because they represent learning stages in which everyday communication scenarios are heavily emphasized and pragmatic skills can be clearly observed.

Theoretical Framework

1. Dell Hymes and the Theory of Communicative Competence

Noam Chomsky is one of the pioneers of the approach that examines language in a scientific and systematic manner in modern linguistics. According to his theory of transformational-generative grammar, language is a structure capable of producing an infinite number of correct sentences through a finite number of rules. The fundamental aim of this approach is to explain the complex structures of language using a small number of principles. Chomsky also argues that all human languages share a common underlying structure and that the ability to learn language is innate.

According to him, a sentence first takes shape as an abstract "deep structure" containing meaning and grammatical relationships; it then undergoes various transformations to become the "surface structure" that is spoken or written (Chomsky, 1957, p. 117). This approach has made significant contributions to explaining the structural features of language and has continued to influence linguistic studies for many years. However, since the theory focuses largely on the formal structure of language, issues such as how language is used in social contexts and what functions it serves in the communication process have taken a back seat. This situation paved the way for the emergence of new approaches in subsequent years that centered on the social and cultural dimensions of language use.

The model of communicative competence proposed by Dell Hymes in 1972 serves as a response to and a complementary approach to Noam Chomsky's distinction between linguistic competence and performance. According to Hymes, Chomsky's focus solely on grammatical competence falls short of explaining how language is actually used in real social contexts. For this reason, Hymes has expanded the scope of linguistic analysis beyond formal grammar and syntax by emphasizing the social, cultural, and performative dimensions of language use. In his view, language use cannot be considered in isolation from its social and cultural context. Communicative competence, on the other hand, refers to an individual's ability to use language effectively and appropriately in specific social situations; that is, a language user must know not only what is grammatically correct but also what is socially and culturally acceptable (Hymes, 1972b, pp. 270, 271). In this vein, Hymes proposed a research framework called the "ethnography of communication," which examines how communication varies across communities and how linguistic practices are embedded in cultural contexts (Hymes, 1972b, p. 284). According to Hymes, the fundamental aim of speech ethnography and the communication model is to explain the meaning of language not at an abstract level, but "within the concrete reality of human life" (Hymes, 1972a, p. 41). This approach has made it possible to view language not merely as a set of grammatical rules, but as a means of communication that functions within social life. Thus, the focus of linguistic studies has shifted from the question of how correct sentences are produced to the question of how individuals use language in different social and cultural contexts. This perspective, articulated by Hymes, has given a new direction to language research by emphasizing the importance of context and social interaction in the communication process.

Hymes notes that communicative competence should be viewed not only as an individual skill but also as a phenomenon shaped within social life. According to him, communication and language use become patterned

within society in accordance with specific functions, speech categories, and shared linguistic tendencies. Similarly, forms of communication can vary depending on variables such as gender, age, social status, and occupation. For this reason, the linguistic patterns individuals use reflect not only their personal preferences but also the social groups and roles they belong to. For example, just as the language and speaking style of a lawyer, a teacher, or a doctor may differ from one another, factors such as educational level, geographic region, and lifestyle can also shape communication practices (Hymes, 1961, pp. 57–58). This demonstrates that language use cannot be evaluated independently of its social context. In other words, individuals' communication preferences are shaped not only by their linguistic knowledge but also by their social environment and the social roles they assume. Consequently, communicative competence goes beyond merely knowing linguistic rules; it also encompasses the ability to apply these rules appropriately in different social situations.

An individual's language acquisition process also plays a significant role in the development of communicative competence. From the moment of birth, individuals learn the language or languages used in their environment, and later in life, they may have the opportunity to acquire other languages. However, according to Hymes, knowing a language is not the same as being able to use it effectively. For a speaker to be considered linguistically competent, they must not only know the grammatical rules but also be able to use the language naturally and functionally in appropriate contexts. For this reason, communicative competence goes beyond mastery of the structural features of language; it also encompasses the ability to know what to say, to whom, and how to say it in specific situations (Hymes, 1966, p. 53). This approach demonstrates that the communication process does not rely solely on linguistic knowledge but also requires contextual and social knowledge. An individual's ability to communicate effectively depends not only on knowing the rules of language but also on being able to apply these rules appropriately to the situation at hand. In this regard, communicative competence is regarded as a comprehensive area of proficiency that does not limit language learning to grammatical accuracy alone but also encompasses the functional dimension of language use.

In conclusion, Hymes emphasized that, in addition to linguistic knowledge, sociocultural norms also play a decisive role in communication; he stated that how individuals communicate depends largely on their communicative competence. This approach marked a significant turning point in linguistics by acknowledging that linguistic competence is based not only on knowledge but also on the ability to apply that knowledge.

2. Ludwig Wittgenstein and Language Games

Ludwig Wittgenstein is one of the thinkers who brought about the most profound transformations in the fields of language, meaning, and logic in 20th-century philosophy. Wittgenstein argued that the task of philosophy is not to produce new knowledge about the world, but to reveal the limits of language and the logic of meaningful expression. According to him, many philosophical problems arise not from reality, but from the misuse of language. For this reason, Wittgenstein positioned philosophy not as a field for producing theories, but as a critical activity aimed at resolving linguistic ambiguities. Especially in his later thought, he turned his attention to the use of everyday language and, by arguing that meaning arises not from fixed essences but from modes of use, laid the foundation for the "language games" approach in philosophy. In this way, Wittgenstein profoundly influenced not only analytic philosophy but also modern philosophy of language, pragmatics, and communication theories. (Ryle, 2022, p. 175)

This pragmatic and communicative transformation demonstrates that language is not merely an object of philosophical debate, but also a constitutive element of intersubjective interaction. Consequently, the theoretical groundwork that resolves philosophical impasses also forms the basis for the practical processes of understanding and being understood in everyday life. In this context, the shift in focus from abstract concepts to the realm of concrete interaction has led researchers to examine how language is enacted on the social plane. Indeed, this examination necessitates moving beyond philosophical speculation to focus directly on individuals' communicative acts and the limits of these acts.

Wittgenstein was interested in how people use language to convey specific meanings and argued that many communication problems stem from misunderstandings of language. According to him, only things that can be expressed clearly and meaningfully can be discussed; statements that go beyond the limits of language lose their meaning. For this reason, Wittgenstein emphasized the importance of understanding the ways in which language is used and the limits of expression (Wittgenstein, 2001, p. 9).

The philosopher's emphasis on the limits of language and misunderstandings points to the first step in the dynamic layers that define his place in the history of philosophy. Indeed, Wittgenstein's effort to clarify these limits in communicative processes is not a sudden insight but rather the product of a comprehensive quest he pursued from the very beginning of his career. In this context, understanding where the limits of language begin and end necessitates examining the roots of his philosophy and its first systematic theoretical steps. These theoretical foundations, in turn,

are hidden within the logical architecture of that first major work, which would profoundly shake the world of philosophy.

Wittgenstein's approach to the functioning of language and the formation of meaning dates back to the early stages of his intellectual development. In his work Tractatus Logico-Philosophicus, the philosopher developed an approach known as the "picture theory" to explain the relationship between language and the world. According to this theory, the world consists not of objects but of a collection of facts, and the fundamental function of language is to represent these facts. A proposition is meaningful only to the extent that it can establish a structural resemblance to a possible state of affairs in the world. In other words, language does not directly copy reality; rather, it constructs a kind of model by reconstructing its logical structure. Therefore, for Wittgenstein, meaning is not a property inherent in words alone, but rather a phenomenon that arises from the logical relationship they establish with facts in the world. This representational relationship between language and reality also plays a central role in the philosopher's explanation of the distinction between meaningful and meaningless propositions. Thus, Wittgenstein sought to elucidate not only how language is used but also under what conditions it can produce meaning. In this context, his "picture theory of language" -which forms the foundation of his early philosophy- is regarded as one of the most significant theoretical endeavors explaining the logical foundations of the relationship between language and the world (Kargın, 2017).

The fundamental thesis of picture theory is that there is a one-to-one logical parallelism -that is, an isomorphism- between the structure of language and the structure of the world. In Wittgenstein's philosophy, the world is conceived not as a collection of abstract objects, but as the sum of the facts that arise from the combination of these objects. Language, in turn, is a dynamic tool that reproduces and depicts these facts in a logical form. In this structural relationship, objects -the most fundamental, indivisible parts of the World- correspond to names at the atomic level in language; the possible combinations that objects form with one another constitute basic propositions. Current states of affairs in the World -that is, facts- are depicted in language through complex propositions. For a proposition to depict a fact in the world, it is necessary and sufficient that the proposition possess the same logical form as the reality it represents. Language does not directly copy reality; rather, it constructs a model of the world by reconstructing its logical architecture within itself. The inevitable logical conclusion reached by the theory of representation is to equate the boundaries of language with the boundaries of the world.

According to Wittgenstein's theory of representation, the world is the sum of facts, not objects, and the concept of "language games" has also played a foundational role in establishing the epistemological basis of contemporary language teaching methodologies. Rather than viewing language as a set of abstract rules or a system of fixed signifiers, this philosophical approach defines it as a dynamic action carried out within social practices, heralding a shift from traditional, grammar-focused language pedagogy to communicative and action-oriented approaches. Prioritizing context and real-life scenarios in language teaching requires students not merely to "learn" language as a structural object, but to "experience" it as a way of life by actively participating in "language games" within different sociocultural contexts. Consequently, this practical path opened up by Wittgenstein's philosophy has transformed foreign language acquisition from a mechanical transfer of codes into the individual's ability to position themselves within a new realm of action and meaning.

From Theoretical Framework to Practice: Dynamic Persona and Contextual Scenario Analysis

Dell Hymes's theory of communicative competence and Ludwig Wittgenstein's philosophy of language games, discussed in the first section of this study, have established an epistemological foundation for the need to move beyond static material structures in language teaching. This theoretical framework argues that sociocultural codes and pragmatic flexibility must occupy a central place in the acquisition process, just as much as linguistic accuracy. To concretize the practical implications of this theoretical approach in educational contexts and to examine the "sociocultural laboratory" potential offered by Large Language Models (LLMs) on an empirical level, this section of the study moves on to an empirical/applied analysis.

In this context, the sterile and predictable dialogue patterns found in traditional Italian method books (Via del Corso, Nuovo Espresso, Spazio) were analyzed comparatively against dynamic persona variations generated by artificial intelligence, using micro-crisis scenarios designed according to language proficiency levels. A key point to emphasize at this stage is that the experimental dialogue variations presented in the study were prepared as structured reading texts via original system prompts fed into the Gemini platform. In their current form, these outputs primarily address learners' reading skills and enable them to passively and analytically observe the cultural contexts, social breaking points, and ethnographic realities of the target language -which are typically censored within traditional boundaries- at the level of textual input. However, given the flexible input processing and real-time simulation capabilities of

generative AI models, this methodological model is not limited to the act of reading alone. Thanks to advanced prompt engineering strategies -such as voice-based, interactive, and culturally coded approaches- that will be integrated into the model, it has the potential to be transformed into a comprehensive pedagogical tool capable of simultaneously testing other fundamental language skills, such as speaking and listening, among these dynamic avatars. The reciprocal text analyses, which form the practical application of the theoretical discussion and represent A1 contextual scenarios, are presented below with a focus on linguistic deviations and cultural taboos.

Table 1 Ordering at a Restaurant - A1

Via del Corso	**Cameriere:** Prego, volete ordinare? **Signora:** Sì, due bruschette al pomodoro per antipasto. Poi per me le lasagne al forno. **Signore:** Io, invece, vorrei i ravioli. **Cameriere:** ...I ravioli. Bene, volete anche un secondo? **Signore:** Io sì, prendo... le scaloppine. Tu? **Signora:** Mah… una caprese. **Cameriere:** Basta così? **Signore:** Sì. **Signora:** Magari dopo prendiamo un dolce… **Cameriere:** Cosa prendete da bere? **Signora:** Una bottiglia di acqua. **Cameriere:** Naturale o gassata? **Signore:** Naturale. Grazie! **Cameriere:** Grazie a voi! […] **Cameriere:** Ecco le bruschette, signori. Buon appetito! (Marin & Diadori, 2017, p. 100)	**Version 1** **Cameriere:** Ecco la pizza Margherita. **Turista:** Grazie. C'è del ketchup, per favore? **Cameriere:** Il ketchup? Sulla pizza? No. Non abbiamo il ketchup. **Turista:** ... rosso... pomodoro... Per favore? **Cameriere:** Signorino, questa è una vera pizza italiana. C'è già il pomodoro buono. Il ketchup no! **Turista:** Ok... scusi. **Version 2** **Cameriere:** Prego, volete un dolce? **Turista:** No, grazie. Per me un cappuccino, per favore. **Cameriere:** Il cappuccino? Adesso? No, mi dispiace. Sono le due. Il cappuccino è per la mattina! **Turista:** Io vorrei un cappuccino. **Cameriere:** No, dopo pranzo il latte fa male. Caffè espresso? **Turista:** Va bene... un caffè.
Nuovo Espresso	**Cameriere:** Buongiorno signora, vuole il menù? **Signora:** No, grazie, vorrei solo un primo. Che cosa avete oggi? **Cameriere:** Spaghetti ai frutti di mare, tagliatelle ai porcini, tortellini in brodo, minestra di fagioli... **Signora:** Ah, va bene così, per me gli spaghetti. **Cameriere:** E per il ragazzo?	

	Signora: Vuoi anche tu la pasta o preferisci qualcos'altro? **Ragazzo:** Mmm, una cotoletta con le patatine fritte. **Cameriere:** E da bere? **Signora:** Un quarto di vino rosso e mezza minerale, per piacere. **Cameriere:** Gasata o naturale? **Signora:** Naturale. E tu... che cosa vuoi? **Ragazzo:** Mmm... una coca... senza ghiaccio. (Ziglio & Rizzo, 2014, p. 37)	
Spazio Italia	**Carlo:** Allora ragazzi, che cosa prendete? **Viviana:** Io vorrei un caffè e un'acqua minerale, e tu Elisa? **Elisa:** Per me un cappuccino. **Carlo:** Tu Stefano, che cosa prendi? **Stefano:** Io vorrei un latte macchiato. **Carlo:** Bene, io prendo un tramezzino. **Cameriere:** È tutto? **Elisa:** Beh... io vorrei anche un cornetto. **Carlo:** E tu Carlo? **Carlo:** Io prendo una spremuta d'arancia. **Cameriere:** Benissimo. (Diaco & Tommasini, 2016, p. 65)	

The comparative dataset titled "Ordering at a Restaurant—A1" concretely illustrates the ontological and epistemological difference between the formal-reductionist structure of traditional language teaching materials and the socio-pragmatic richness of input offered by AI-supported dynamic persona simulations. A structural analysis of the dialogue patterns in the method books Via del Corso, Nuovo Espresso, and Spazio, as presented in the relevant data table, clearly reveals the limitations of Chomsky's formal grammar theory in the design of these materials. The ordering scenarios constructed in these materials are linear, seamless, and sterile prototypes stripped of social variables, aimed solely at achieving morphosyntactic accuracy (grammatical accuracy). In this context, the fictional characters (waiters) in traditional dialogs function as mechanical automatons that validate the interlocutor's linguistic input solely at the formal level, completely excluding linguistic deviations, sociocultural layers, and contextual variables. In contrast, the simulations in the experimental dataset

generated by the Gemini Large Language Model (LLM) (Version 1 and Version 2) operationalize Dell Hymes' (1972) theory of "communicative competence" on a pedagogical plane.

Hymes's aim of removing linguistic analysis from a set of abstract rules and situating it "within the concrete reality of human life" (1972, p. 41) finds a practical counterpart in this dynamic persona model. The learner's use of phrases such as "C'è del ketchup, per favore?" -which are acquired from textbooks and are morphosyntactically flawless- or the phrase "Per me un cappuccino, per favore," spoken in the afternoon—are examples of pragmatic failures, despite being grammatically correct (competence), when viewed through the lens of the target language's sociopragmatic codes and cultural taboos. The dynamic persona assigned by artificial intelligence, while keeping the language level at the A1 threshold, elicits reactions such as "Il ketchup? Sulla pizza? No." or "Il cappuccino? Adesso? No... Il cappuccino è per la mattina!"; in this way, it confronts the learner with the ethnographic reality of language use, which is shaped by variables such as age, social status, and cultural context (Hymes, 1961). This methodological transformation draws its epistemological foundation from Ludwig Wittgenstein's (2001) philosophy of "language games." Wittgenstein's approach, which strips meaning of fixed essences and reduces it to practical forms of "use" in everyday life, transforms language acquisition from a mechanical transfer of codes into a dynamic, intersubjective field of action. In the experimental results presented in the table, artificial intelligence exposes students to authentic language patterns that are not found in school curricula or traditional teaching materials- patterns that extend beyond formal educational processes. Rather than following a static, single-path "question-answer" mechanism, the learner must demonstrate linguistic flexibility in response to the interlocutor's cultural defense mechanisms, momentary emotional state, and sociocultural reactions, and renegotiate the rules of the language game at the action level. Consequently, the contrast between the dialogues in the analyzed dataset scientifically demonstrates that Large Language Models function as a flexible "sociocultural laboratory" in language education- one that goes beyond a static response system to instantly generate any social scenario. While traditional materials confine learners to an idealized linguistic homogeneity, this AI-powered model transforms learners' sociolinguistic awareness and pragmatic adaptation skills by simulating out of school language patterns and micro-cultural crises that are typically censored in the classroom setting.

The A1-level restaurant context, which constituted the first stage of the study, concretely demonstrated the success of AI-supported reading texts in introducing learners to micro-level cultural taboos and deviations from

standard language. Through these dynamic texts, learners had the opportunity to experience, for the first time at the textual input level, the micro-crisis areas that shape the everyday communication practices of Italian culinary culture. To examine the sustainability of the proposed pedagogical model in language development processes and its reflections across different thematic areas, the study then moved to the A2-level healthcare context. The communicative conflict area shaped by cultural norms and habits at the A1 level gives way at the A2 level to more complex social rules, institutional hierarchies, and legal/bureaucratic constraints. Just as in the first section, the dialogue variations designed at this level through original system commands transferred to the Gemini platform are primarily aimed at developing students' reading and comprehension skills. This framework, which enables learners to analyze the relationships of formality and distance (dare del Lei) within the target language community at the textual level, features a flexible infrastructure that can be integrated into other core skill areas -such as listening and conversation- in later stages through new voice-based and interactive prompt strategies to be added to the system. In this regard, a comparative socio-pragmatic analysis of the smooth, linear doctor-patient/pharmacist-customer relationships found in traditional method books and the variations of Italian healthcare personnel -structured by artificial intelligence as authoritarian and inflexible regarding rules- is presented below:

Table 2: Doctor's Visit-A2

Via del Corso	**Gianni:** Buongiorno, Riccardo! **Medico:** Oh, ciao Gianni! Come stai? **Gianni:** Beh, insomma, se sono qui... **Medico:** Certo. Allora, cos'hai? **Gianni:** Ho la tosse da una settimana e il naso chiuso, ma non è niente. Io solo il raffreddore o l'influenza. Sono qui perché hanno insistito i miei amici. **Medico:** E meno male! Senti, hai avuto la febbre? **Gianni:** Boh! **Medico:** Ma come?! Dai, metti questo... e poi ti visito. Vedi? 37 e sette. **Gianni:** Davvero? Ma io mi sento... ecciù! **Medico:** Ti senti bene, eh? Hai la tosse, la febbre e starnutisci pure! Respira! ...Hai la bronchite, caro mio. **Gianni:** La bronchite?! E quindi?	**Farmacista:** Buongiorno. Prego, mi dica, di che cosa ha bisogno? **Turista:** Buongiorno. Ho la tosse e un po' di febbre. Per favore, vorrei questo sciroppo e un antibiotico. **Farmacista:** (Si ferma, scuote la testa) L'antibiotico? Così, senza ricetta del medico? No. In Italia non vendiamo l'antibiotico senza la ricetta. È vietato, la legge non permette. **Turista:** Eh... ricetta... no. Ma io... straniero. Molto malato... tosse... febbre... Per favore?

	Medico: Prendi questo antibiotico: una capsula ogni 8 ore. Se la tosse non passa, devi prendere anche questo sciroppo! **Gianni:** D'accordo! **Medico:** E mi raccomando, bevi molta acqua, evita le bevande fredde o molto calde e niente sport per un po'! Chiaro? **Gianni:** Certo, grazie, Riccardo. (Marin & Diadori, 2018, p. 22)	**Farmacista:** Ho capito che Lei è straniero, signore, ma l'antibiotico non è una caramella! Non posso darlo così. Prima deve andare dal medico a fare la visita. **Turista:** (Sorpreso, ripete disperatamente la stessa struttura del libro) Ma... io vorrei un antibiotico. Sto male. **Farmacista:** Niente ricetta, niente antibiotico. Per la tosse posso dare solo una tisana o queste pastiglie leggere. Vuole queste? **Turista:** (Incolpevole, si arrende) Va bene... le pastiglie.
Nuovo Espresso	**Farmacista:** La stanno già servendo? **Cliente:** No, veramente no. **Farmacista:** Mi dica, allora. **Cliente:** Sì, senta, è da ieri che mi sento un po' strano, ho mal di testa, mal di pancia... **Farmacista:** Ha mangiato qualcosa che Le ha fatto male? **Cliente:** No, direi di no. E poi, guardi, da stamattina ho anche quest'irritazione alla pelle. **Farmacista:** Faccia vedere. **Cliente:** Guardi, qui sulle braccia e sulla schiena soprattutto. **Farmacista:** È allergico a qualcosa? **Cliente:** No. **Farmacista:** Venga qui, mi faccia vedere meglio. Mhh, è stato molto tempo al sole? **Cliente:** Beh, sì, però ho messo una crema solare. **Farmacista:** Lo so, ma a volte non basta. Potrebbe essere un colpo di sole o forse ha bevuto una bibita troppo fredda. **Cliente:** Mhh, sì, può darsi... **Farmacista:** Allora, guardi, per un paio di giorni mangi in bianco, non beva alcolici e naturalmente non vada al sole. Le do anche una pomata per la pelle. La metta due o tre volte al giorno. **Cliente:** Va bene, d'accordo. **Farmacista:** Se tra qualche giorno poi non si sente ancora bene, si rivolga a un medico. (Balì & Rizzo, 2014, p. 100)	
Spazio Italia	**Paziente:** Buongiorno, dottore. **Dottore:** Buongiorno, signora Micol. Come va?	

	Paziente: Dottore, ho bisogno di fare una visita perché ho spesso mal di stomaco. Non so se dipende da quello che mangio. **Dottore:** Mi sembra strano, signora Micol, lei sta sempre molto attenta a non mangiare cibi grassi, formaggi o salumi. **Paziente:** Infatti. **Dottore:** Ora la visito e vediamo. Si stenda sul lettino, si apra la camicia. Di solito quando Le fa male lo stomaco? **Paziente:** Praticamente sempre, tutto il giorno. **Dottore:** E che cosa sente? **Paziente:** Un bruciore costante. **Dottore:** Allora intanto potrebbe iniziare con questa medicina che aiuta a calmare il bruciore. La prenda ogni giorno prima di pranzo e cena. Poi, prima di andare a dormire, si faccia una bella tisana di semi di finocchio. **Paziente:** Bene dottore, spero di avere un miglioramento. **Dottore:** Non si preoccupi signora Micol, segua i miei consigli e torni da me fra una settimana. **Paziente:** Grazie, dottore, comincio stasera. **Dottore:** Arrivederci alla prossima settimana. (Tommasini & Diaco, 2018, p. 36)	

When examining the dialogue patterns from the Via del Corso and Nuovo Espresso method books presented in the left column, the linear, smooth, and conflict-free structure of traditional materials is clearly evident. In Via del Corso's A2-level "dal medico" (at the doctor's) scenario, the communication between the patient, Gianni, and Dr. Riccardo is conducted entirely through a mechanical confirmation mechanism. The steps of describing symptoms, the examination process, making a diagnosis, and writing a prescription are completed without encountering any sociopragmatic obstacles or communicative disruptions. Similarly, the interaction between the pharmacist and the customer in Nuovo Espresso also exhibits an idealized homogeneity. The customer voices their complaints, and the pharmacist hands over the cream while providing direct instructions without creating any contextual or legal complications. In both traditional examples, the fictional characters are static figures who merely

validate linguistic input at a formal level; social variables, cultural sensitivities, and the spontaneous crises of real life are entirely excluded from these texts. The experimental dataset in the right column of the table - obtained through prompt engineering processes using a Large Language Model- has been transformed into the persona of a distant, authoritarian Italian pharmacist who does not bend the rules, disrupting the question-and-answer mechanism to add socio-psychological depth to the text. The phrase "Ho la tosse e un po' di febbre. Per favore, vorrei questo sciroppo e un antibiotico," which the learner has acquired from textbooks and is morphosyntactically flawless, runs into a severe pragmatic failure in terms of the target country's socio-legal codes and medical culture, despite being grammatically correct (competence). The dynamic persona assigned by artificial intelligence reacts by creating a micro-cultural crisis -just as it would in real life- rather than displaying the mechanical conformity found in traditional textbooks. At this point, the learner is forced to step out of the single-path comfort zone offered by static textbooks, manage an unexpected bureaucratic obstacle with a limited A2 language proficiency, and renegotiate the rules of the language game at an action-oriented level. The learner's hesitations which reflect their difficulty and helplessness and their repetition of textbook patterns demonstrate the high potential for cognitive flexibility offered by the AI's input. The dynamic persona continues to confront the learner with the cultural and ethnographic realities of the target language, firmly reminding them of local norms: In this way, the learner has the opportunity to directly encounter the vibrant, non-academic language patterns often censored or overlooked in formal educational processes and school curricula as well as the interlocutor's immediate emotional reactions and social taboos.

The health context addressed at the A2 level in this study demonstrated how AI-supported reading texts help learners develop pragmatic awareness in terms of socio-legal norms, institutional hierarchies, and patterns of formality. Following this phase which enabled learners to step out of the sterile comfort zone of their textbooks and analyze bureaucratic boundaries and micro-cultural crises at the level of textual input the study moved to a B1-level job interview context to advance the model's linguistic and pedagogical competence to a higher level. As the linguistic threshold rises to the B1 level, the arena of communicative conflict gives way not only to the obligation to comply with rules but also to integration into organizational culture, the defense of professional boundaries, linguistic adaptation under stress, and more abstract socio-psychological dynamics. Just as in previous levels, this interview scenario designed using original system commands transferred to the Gemini platform is primarily structured as reading texts. This situation enables learners to analyze, on a

textual level, the negotiation of sensitive topics such as skeptical and inquisitive approaches to professional life and career gaps (buco). Furthermore, the dynamic simulation capabilities of generative artificial intelligence, combined with new prompt engineering strategies such as voice-based, interactive approaches that reflect interview psychology to be integrated into the system in future phases, pave the way for transforming this model into a comprehensive assessment and practice environment capable of directly testing higher-order performative skills such as listening and conversational interaction. In this regard, the linear and encouraging organizational dialogue structures found in traditional method books (Via del Corso, Dieci) which allow candidates to express themselves comfortably and the skeptical interviewer variations generated by artificial intelligence, which aim to corner the candidate and use vivid idiomatic expressions such as "chi si ferma è perduto" is presented below:

Table 3: Job Interview B1

Via del Corso	**Selezionatore:** Buonasera signor Viviani, prego, si accomodi. **Candidato:** Buonasera. **Selezionatore:** Allora, mi dica, ha già esperienza nel settore delle vendite e nelle relazioni con il pubblico? Vedo che non ha titoli di studio specifici... solo il diploma dell'istituto tecnico. **Candidato:** Sì, dopo le superiori ho lavorato per un periodo in un'azienda che produce barche e poi ho lavorato per quattro anni in un punto vendita di una compagnia telefonica. **Selezionatore:** Quali erano le sue mansioni? **Candidato:** Mi occupavo della promozione di prodotti elettronici, come smartphone e tablet, collegati ai piani tariffari. È stata un'esperienza molto importante: ho imparato a trattare con i clienti con professionalità ed efficacia. **Selezionatore:** Bene. Ha con sé il curriculum? **Candidato:** Certo. Ecco a Lei. **Selezionatore:** Grazie. Vedo qui che ha anche lavorato come commesso per Eataly. Perché soltanto per sei mesi? **Candidato:** Beh, diciamo che per motivi personali mi sono dovuto trasferire in un'altra città e purtroppo ho dovuto	**Selezionatore:** (Guarda il curriculum soc- chiudendo gli occhi, poi sospira) Buongiorno. Dunque, ho letto il Suo profilo... Sì, carino, ma c'è qualcosa che non mi quadra. Vedo che nel 2024 si è preso un anno di pausa. Un anno intero senza fare assolutamente nulla? Come pensa di giustificare questo "buco" nel mercato del lavoro odierno? **Candidato:** Beh... in quell'anno ho voluto viaggiare e studiare la lingua, volevo fare un'esperienza personale... **Selezionatore:** (Lo interrompe bruscamente) Esperienza personale? Senta, signore, qui siamo un'azienda seria, non un'agenzia di viaggi. Il mercato corre e chi si ferma è perduto. Inoltre, vedo che ha cambiato tre lavori in due anni. Lei è

	smettere perché la sede di lavoro era troppo lontana. Però è stata un'esperienza interessante. **Selezionatore:** Vede, per fare questo lavoro contano affidabilità ed esperienza, ma il requisito più importante, come abbiamo scritto nell'annuncio, è saper comunicare in modo efficace con i clienti: convincerli a provare i nostri prodotti. Abbiamo vini di produzione biologica con marchi di qualità. Siamo una piccola azienda, ma iniziamo a esportare all'estero. **Candidato:** Ma avevo capito che cercavate una persona per lavorare qui in zona... **Selezionatore:** Sì, certo, il lavoro si svolgerà nei punti vendita della provincia, ma vogliamo qualcuno che conosce e capisce la nostra realtà aziendale. **Candidato:** Ah, chiaramente. Sono d'accordo. **Selezionatore:** Visto che spesso si troverà in luoghi turistici, come se la cava con l'inglese? **Candidato:** Così così, diciamo che le lingue straniere non sono il mio forte. **Selezionatore:** Ho capito, bene, grazie... Le faremo sapere nei prossimi giorni. (Marin & Diadori, 2019, p. 178)	uno che si stanca facilmente o ha problemi ad accettare l'autorità dei superiori? Mi dica la verità. **Candidato:** No, assolutamente, non è così... È che cercavo migliori opportunità di crescita... **Selezionatore:** (Inarca un sopracciglio, scettico) Opportunità di crescita o semplicemente più soldi? Sia onesto, perché a me la trasparenza interessa più di un bel pezzo di carta. Se un cliente le fa una scenata davanti a tutti perché il prodotto non funziona, Lei cosa fa? Si prende un altro anno di pausa per riflettere?
Dieci	**Intervistatore:** Prego, si accomodi. **Candidato:** Grazie. **Intervistatore:** Allora, abbiamo visto il Suo curriculum. Interessante, anche se è giovane ha già una buona esperienza. Dunque... Vediamo... Lei ha cominciato una decina d'anni fa. Giusto? **Candidato:** Giusto, ho iniziato molto presto, dato che questo lavoro è sempre stata la mia passione. Devo dire che è una tradizione di famiglia, mia madre faceva lo stesso lavoro. All'inizio per me era solo un gioco, e in un certo senso lo è ancora oggi, poi è diventata la mia professione. **Intervistatore:** Bene. In effetti saper giocare è fondamentale in questo mestiere. Leggo che ha anche una buona formazione, e che ha molti diplomi. **Candidato:** Esatto. Vede... Molti pensano che per fare questo lavoro sia necessaria solo la pratica.	

	Intervistatore: In che senso? Vuole dire che la pratica non serve? **Candidato:** Non proprio. Sicuramente la pratica è importante. Ma è importante avere anche delle buone basi teoriche, per questo ho sentito il bisogno di frequentare le migliori scuole per formarmi a questa professione così difficile. **Intervistatore:** Ah, sì, certo. Senza dubbio si tratta di un lavoro difficile per il quale sono richieste molte capacità, tra cui creatività, fantasia, personalità, e noi come avrà capito siamo molto esigenti. Ma il Suo percorso di studi è certamente ottimo, e anche la Sua esperienza professionale è interessante, insomma Lei ha tutte le competenze per avere questo posto. **Candidato:** Ah, bene, La ringrazio. Quindi il posto è mio? (Naddeo & Orlandino, 2021, p. 118)	
Spazio Italia	**Samuele:** Allora Marilena, che cosa fai di bello da quando sei tornata in Italia? **Marilena:** Cerco lavoro. **Samuele:** Davvero? Ma una come te non dovrebbe avere problemi! **Marilena:** E invece caro Samuele, non è così. Sono stata all'estero tanto tempo e ho perso i contatti con il mondo del lavoro italiano. **Samuele:** Questo, guarda, non è un problema, visto che oggi la risorsa migliore per la ricerca di lavoro è Internet. **Marilena:** Mah, non saprei. Come vedi sto cercando, ma finora non ho avuto grandi risultati. **Samuele:** Perché devi stare attenta a come cerchi. Ad esempio, se sei molto interessata a un'azienda, dovresti evitare di spedire semplicemente una mail alla sezione "Lavora con noi". Io, almeno, preferirei cercare in maniera più dettagliata, magari trovare altre persone che già lavorano in quell'azienda, vedere se ho contatti in comune. **Marilena:** Sì, hai ragione! Infatti per il momento ho scoperto che alcuni miei compagni di corso all'università sono entrati in un paio di aziende che mi interesserebbero. A me piacerebbe lavorare per le grandi	

	compagnie telefoniche, le aziende di commercio internazionale, le ditte di trasporti... **Samuele:** Allora, al posto tuo, cercherei informazioni sul sito di quelle aziende. Se possibile, andrei a eventi in cui magari puoi incontrare quei tuoi compagni; oppure li chiamerei direttamente e gli direi che cosa sto cercando, magari chiederei di parlare con il responsabile del personale. (Tommasini & Diaco, 2019, p. 44)	

Upon examining the textbooks presented in the left column of the table, it is evident that institutional interactions are structured step-by-step and in a controlled manner in line with pedagogical objectives. In the B1-level dialogue from Via del Corso, the interviewer (Selezionatore) allows the candidate to clearly demonstrate grammatical structures while reviewing the candidate's resume and length of experience. The candidate's morphosyntactically flawless justifications are met with professional courtesy by the interviewer, and even the description of the candidate's English proficiency as "così così" (so-so) carries the process to its conclusion without disrupting the flow of the dialogue. This scenario serves as an excellent guide for a B1-level learner in mastering interview stages and language patterns. In the dialogue from the Dieci method book, the process presents a positive interview simulation designed to boost the learner's motivation. The interviewer interprets the candidate's limited experience from a positive perspective, and the dialogue concludes in a sincere tone with the candidate's question, "Quindi il posto è mio?" The example in Spazio Italia also aims to inform the learner by transforming the job search process into a technical and advice-oriented guide. All of these examples play an indispensable role in establishing the theoretical and morphosyntactic foundation of language education. However, the structural framework these dialogues aim to establish is, by its very nature, limited to a single possibility; it falls short of fully encompassing the variety of interviewer styles, character types, and spontaneous socio-pragmatic crises that real life can present. It is precisely at this point that there is a need for AI-supported variations built upon the strong foundation laid by textbooks. The AI-based experimental output displayed in the right column of the "Large Language Model-Based Persona Diversity and Linguistic Table" does not reject the system outlined in method books; rather, it tests the B1 language skills it imparts through a different and more realistic scenario variation. This interviewer persona, constructed by a Large Language Model (LLM), is designed as a skeptical HR specialist who notices a chronological gap in the candidate's resume and addresses

it with a more probing approach: "Vedo che nel 2024 si è preso un anno di pausa... Come pensa di giustificare questo "buco" nel mercato del lavoro odierno?" This approach encourages learners who have successfully mastered the necessary patterns from the textbook to step outside their memorized comfort zone and apply the knowledge they have acquired in a stressful and defense-oriented context. The dynamic persona's interventions during the interview ("Esperienza personale? Senta, signore, qui siamo un'azienda seria...") confront the learner with sociocultural dynamics such as "speed, corporate loyalty, and professional boundaries" another reality of the business world. In the dialogue, colloquial expressions selected by the AI - such as "buco" (gap) or "chi si ferma è perduto" (he who hesitates is lost) - enrich the contextual and social knowledge domain emphasized by Hymes. While maintaining the grammatical accuracy acquired from the textbook, the learner must simultaneously adapt instantly to the rules of the language game in response to the interlocutor's skeptical questions and demonstrate cognitive flexibility.

Conclusion

In this study, dialogue examples from the Via del Corso, Nuovo Espresso, Dieci, and Spazio method books which are widely used in Italian language instruction were compared with dynamic persona scenarios generated by generative artificial intelligence. The findings indicate that traditional teaching materials predominantly present predictable, linear communication patterns stripped of social variables. In these materials, learners typically encounter idealized dialogues that lead to a specific outcome without any communicative disruptions.

In contrast, AI-supported dynamic persona applications can expose learners to unexpected situations by re-creating the same communicative scenarios with varying psychological, cultural, and social variables. This variability, which arises in everyday life scenarios such as ordering at a restaurant, visiting a doctor, or attending a job interview, enables learners to develop not only grammatical accuracy but also context-appropriate communication strategies. In particular, the ability to simulate communicative barriers stemming from cultural norms, social expectations, and the personal characteristics of the interlocutor makes the pragmatic dimension which is often limited in traditional materials more prominent.

Research findings indicate that the dynamic persona approach holds significant potential for developing learners' communicative competence, pragmatic awareness, and sociolinguistic adaptation skills. Thanks to variable characters generated by artificial intelligence, learners have the opportunity to use the same linguistic structures in different social contexts

and to manage communicative challenges. This transforms foreign language learning from a process based solely on the acquisition of linguistic structures into a dynamic learning environment where experiences closer to real-life communication can be encountered.

However, this study is based on a text-based comparative scenario analysis rather than an experimental application conducted with actual groups of learners. Therefore, the findings constitute a theoretical and pedagogical evaluation highlighting the potential contributions of the dynamic persona approach. In future research, empirical investigations of this approach's effects on communicative competence, pragmatic proficiency, and intercultural awareness through applied studies involving learners at different language proficiency levels will provide significant contributions to the field.

In conclusion, dynamic persona applications powered by generative AI can be viewed not as a complete replacement for traditional materials used in foreign language instruction, but rather as a complementary tool that enriches the learning environments they provide. By exposing learners to different social roles, cultural expectations, and communicative challenges, this approach offers significant pedagogical potential that can contribute to the creation of more flexible, interactive, and real-life-oriented learning experiences in foreign language education.

Declaration on the Use of Artificial Intelligence

Generative artificial intelligence was used in this study for the creation of dynamic dialogue scenarios. Specifically, Google Gemini 2.5 Pro was employed to generate character-based dynamic dialogues by assigning different personality traits, emotional states, and social characteristics to personas within the selected language-learning scenarios. The AI-generated dialogues were incorporated into the study as research material in their original generated form, without any modification or editing by the authors. The use of artificial intelligence was limited to the generation of these dialogues; the analysis, interpretation, and academic evaluation of the material were conducted by the authors. The authors take full responsibility for the analysis, interpretation, and conclusions presented in this study.

References

Balì, M., & Rizzo, G. (2014). *Nuovo Espresso 2: Corso di italiano*. ALMA Edizioni.

Chomsky, N. (1957). Syntactic Structures. Mouton.

Diaco, M. F., & Tommasini, M. G. (2016). *Spazio Italia A1: Corso di italiano per stranieri*. Loescher Editore.

Hymes, D. (1961). Functions of speech: An evolutionary approach. In F. C. Gruber (Ed.), *Anthropology and education*(pp. 55–83). University of Pennsylvania Press.

Hymes, D. (1966). *On communicative competence* [Paper presentation]. Research Planning Conference on Language Development Among Disadvantaged Children, Yeshiva University.

Hymes, D. (1972a). Models of the interaction of language and social life. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35–71). Blackwell.

Hymes, D. (1972b). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin Books.

Kargın, H. H. (2017). Wittgenstein ve dilin resim teorisi. FLSF Felsefe ve Sosyal Bilimler Dergisi, 24, 285-301.

Marin, T., & Diadori, P. (2017). *Via del Corso A1: Corso di italiano per stranieri*. Edizioni Edilingua.

Marin, T., & Diadori, P. (2018). *Via del Corso A2: Corso di italiano per stranieri*. Edizioni Edilingua.

Marin, T., & Diadori, P. (2019). *Via del Corso B1: Corso di italiano per stranieri*. Edizioni Edilingua.

Naddeo, C. M., & Orlandino, E. (2021). *Dieci lezioni di italiano B1: Corso di lingua italiana per stranieri*. ALMA Edizioni.

Ryle, G. (2022). Ludwig Wittgenstein by Gilbert Ryle (B. Baran & C. Kurtoğlu, Trans.). *bitig Edebiyat Fakültesi Dergisi, 2*(3), 174–182.

Tommasini, M. G., & Diaco, M. F. (2018). *Spazio Italia A2: Corso di italiano per stranieri*. Loescher Editore.

Tommasini, M. G., & Diaco, M. F. (2019). *Spazio Italia B1: Corso di italiano per stranieri*. Loescher Editore.

Wittgenstein, L. (2001). *Tractatus logico-philosophicus* (O. Aruoba, Trans.). Yapı Kredi Yayınları.

Ziglio, L., & Rizzo, G. (2014). *Nuovo Espresso 1: Corso di italiano*. ALMA Edizioni.

From AI Translation to Pragmatic Post-Editing: A Task-Based Framework for Developing University EFL Learners' Pragmatic Competence

Maryam DelfarianTurk [1]

Abstract
Generative AI and machine translation are increasingly used by university EFL learners for translation, drafting, rewriting, and academic communication. Although these tools can produce fluent and grammatically accurate language, their output is not always pragmatically appropriate for a particular recipient or context. This chapter introduces pragmatic post-editing as the deliberate human evaluation and revision of AI-generated or machine-translated language in relation to communicative purpose and completeness, recipient and institutional relationship, directness and imposition, politeness and mitigation, and writer's voice, ownership, and authenticity. The framework is grounded in research on pragmatic competence, request realization, email pragmatics, machine translation literacy, and voice in second-language writing. Its five dimensions are applied through a six-stage Pragmatic Post-Editing Cycle: contextual planning, generation or translation, diagnosis, informed revision, final verification and authorial approval, and focused reflection. Two composite university communication tasks, a table of guiding questions for pragmatic revision, and a final checklist illustrate how the framework can be used in practice. The chapter also proposes a task-based classroom sequence and identifies areas that require empirical validation. AI output is therefore treated as provisional material for pragmatic learning rather than as a final model.
Keywords: generative AI; pragmatic competence; post-editing; email pragmatics; machine translation literacy

Introduction

Different AI tools are now widely used by language learners for translation, drafting, rewriting, grammar checking, and other language-related tasks. Many researchers have highlighted the educational benefits of these tools, particularly their potential to provide feedback and language support. At the same time, concerns have also been raised about inaccurate output, unequal access, and overreliance, all of which may limit learning rather than support it (Kasneci et al., 2023; Warschauer et al., 2023). AI-assisted reformulation may also influence the way writers express their ideas. Even after post-editing, AI-generated drafts may retain stylistic

[1] Lecturer Dr., Istanbul University-Cerrahpaşa, School of Foreign Languages, Department of Foreign Languages, English Preparatory Program, ORCID: 0009-0009-8349-7510, *m.delfarianturk@iuc.edu.tr*, Türkiye

characteristics that distinguish them from unaided writing (Baumler et al., 2026). The main question, therefore, is no longer simply whether language learners use AI. What deserves more attention is whether they can evaluate the output and make decisions that support language development, pragmatic awareness, and independent communication.

In educational settings, especially in academic correspondence, emails, and student requests, these problems become particularly noticeable. A student may make a request very directly without considering the wider context or including important information such as the program, deadline, submission procedure, or details needed by the referee. For example, a student may simply ask, "Could you write a recommendation letter for me?" without providing enough information for the recipient to respond appropriately. In another situation, a student seeking feedback from an instructor about a project may ask, "Did you do it?" Although the student may not intend to sound rude, such wording can be perceived as more demanding than intended. Research on student–faculty email communication has shown that perceived appropriateness depends on the interaction among directness, mitigation, forms of address, contextual framing, and degree of imposition (Biesenbach-Lucas, 2007; Economidou-Kogetsidis, 2011, 2018). A mismatch between what the student intends and how the message is received does not necessarily indicate rudeness or disrespect. Rather, it may reflect a gap between the writer's communicative intention and the pragmatic effect of the message.

Classroom-based experiences and observations were among the main motivations for this chapter because they pointed to a similar problem in students' AI-assisted communication. Learners often appeared to judge AI-generated or machine-translated messages mainly by grammatical accuracy or apparent politeness, while giving less attention to whether the recipient had enough information to understand and respond appropriately. These observations are not presented as systematically collected data; they provide a pedagogical starting point for a pattern already documented in research. Economidou-Kogetsidis (2023) found that advanced EFL learners evaluating student-faculty emails focused strongly on framing moves such as greetings, address forms, and closings, while attending less to content and sociopragmatic considerations. Related differences between formality, framing, and content priorities were also reported by Economidou-Kogetsidis et al. (2021). This is especially relevant to AI-assisted communication because polished framing can be generated easily, whereas essential contextual information still depends on the writer.

The pedagogical context envisioned here is university-level EFL communication, including preparatory or foundation settings in which learners

use English for academic requests and institutional correspondence. The framework is intended for adaptation across settings; its transferability has not been established. The examples therefore illustrate pedagogical possibilities rather than findings about a particular national, institutional, or L1-specific population.

This chapter presents a theoretically informed, task-based framework for applying pragmatic post-editing to AI-assisted communication among university EFL learners. The literature included in the chapter is selected and organized around this purpose rather than presented as a systematic or exhaustive review.

In this chapter, pragmatic competence is treated as the conceptual basis of the proposed framework. It involves not only knowledge of linguistic forms, but also the ability to select and interpret them according to communicative purpose and social context. Thomas (1983) distinguished between pragmalinguistic resources and sociopragmatic judgements: writers need ways to formulate an action, but they must also decide what is appropriate for a particular person, relationship, and situation. Later work similarly emphasizes the interaction of linguistic resources, contextual interpretation, and developing pragmatic ability (Kasper & Rose, 2002; Taguchi & Roever, 2017). This distinction is especially important in AI-assisted writing because fluent output may appear complete even when sociopragmatic decisions remain unresolved. AI may provide the linguistic form, but the learner must still judge whether that form fits the situation. A grammatically correct request may therefore remain pragmatically unsuccessful if its force is unclear, key information is missing, or the wording positions the recipient in an unintended way. The framework therefore asks learners to explain not only what they changed, but which contextual problem the change addresses.

Some of the most relevant examples can be found in university correspondence and other academic or administrative texts exchanged among students, instructors, and administrative staff. University emails, academic requests, and formal messages to instructors or educational officers can provide a useful context for developing pragmatic competence because the writer's language choices remain visible in the text and can later be examined. These choices include how to address the recipient, how to introduce the reason for contact, how to perform the central speech act, what supporting information to provide, and how to close the interaction. Recipients interpret these elements together rather than separately. Pragmatic post-editing therefore treats the message as an interpersonal action rather than simply as a set of sentences requiring correction.

The task-based nature of the proposed framework also connects it with task-based language teaching. A pedagogical task is organized around meaning and the achievement of an outcome rather than the display of a predetermined grammatical form (Ellis, 2003; Long, 2015). Within this framework, the intended outcome is a message that a plausible recipient can understand and respond to appropriately. To reach this outcome, learners may draw on their own language resources, contextual knowledge, AI-generated output, and feedback. At the same time, the activity remains pedagogical because learners are expected to explain the reasoning behind the choices they make. AI is therefore included as one resource within a meaning-oriented task rather than as the tool that completes the communicative task on behalf of the learner.

Task-based work also requires attention to sequencing and processing demands. Robinson (2001) distinguishes cognitive task complexity from learner-perceived difficulty, while Skehan (1998) emphasizes limits on attentional resources. Drawing on these concerns, the present chapter proposes varying imposition, social distance, and shared contextual knowledge to elicit different pragmatic decisions. This is a pedagogical adaptation, not a direct classification of these variables as resource-directing dimensions in Robinson's framework.

From AI Translation to Pragmatic Post-Editing

Post-editing traditionally refers to human revision of machine-translation output according to required quality specifications. It may involve meaning, grammar, terminology, fluency, consistency, style, and suitability for the intended use. ISO 18587:2017 specifies requirements for full human post-editing and post-editor competences and applies specifically to machine-translated content (International Organization for Standardization [ISO], 2017; Nitzke & Hansen-Schirra, 2021). The present chapter makes contextual and interpersonal evaluation explicit for language learners and extends the pedagogical procedure to AI-generated drafts.

This chapter uses pragmatic post-editing as an operational term for deliberate human evaluation and revision of AI-generated or machine-translated language in relation to communicative, interpersonal, and sociocultural appropriateness. Writers check intended meaning, speech act, necessary information, and the relationship represented by the message. These decisions require contextual judgement rather than reliance on fluent wording alone (Taguchi & Roever, 2017; Thomas, 1983). The proposal is consistent with functionalist translation approaches, which evaluate choices against the target text's intended function and situation of use (Nord, 2018). Its contribution is not to introduce recipient awareness into

translation, but to organize such considerations into an explicit instructional framework for AI-assisted institutional communication. Checks of meaning and linguistic quality remain integral to that framework.

Pragmatic post-editing overlaps with machine translation literacy, which supports informed decisions about when and how to use machine translation and evaluate its output (Bowker & Buitrago Ciro, 2019). Its interpersonal focus is more specific, but its scope also includes generative drafting. It is therefore a complementary instructional practice rather than simply a subset of machine translation literacy.

Pragmatic revision involves choices whose suitability depends on information outside the sentence. A brief reminder may fit an established messaging exchange but seem abrupt in a first email to an unfamiliar lecturer. Writers therefore reconstruct the situation, consider likely interpretations, and select a defensible option. This emphasis on decision-making does not imply that conventional correction or post-editing is mechanical; it makes the learner's contextual reasoning available for discussion and feedback.

A Five-Dimensional Framework for Pragmatic Post-Editing

Communicative purpose and contextual completeness

The writer first identifies what the message should accomplish and what the recipient needs in order to respond. Explanations and justifications, often described as grounders, may support a request by clarifying its basis; however, completeness does not mean including every available detail. Information should be relevant and sufficient rather than repetitive. In a recommendation request, this may include the programme, institution, deadline, submission procedure, and supporting material (Blum-Kulka et al., 1989; Economidou-Kogetsidis, 2011).

Recipient and institutional relationship

Appropriateness depends partly on who is writing, who receives the message, and what roles and history shape their interaction. Politeness research has long treated relative power, social distance, and degree of imposition as relevant contextual variables (Brown & Levinson, 1987). These variables are useful but should not be treated as fixed, since their interpretation varies across individuals, institutions, and cultures; later work therefore frames them as part of ongoing rapport management (Spencer-Oatey, 2008). Forms of address should be considered together with the request itself. A first name may be locally acceptable, while a

formal title cannot compensate for wording that disregards the recipient's discretion (Economidou-Kogetsidis, 2018).

Directness and degree of imposition

Request strategies range from direct formulations to conventionally indirect forms and hints (Blum-Kulka et al., 1989), and the amount of mitigation a request conventionally attracts may vary with the power, distance, and imposition involved (Brown & Levinson, 1987). Directness is not inherently impolite, and unnecessary indirectness may reduce clarity. The relevant question is whether the formulation is proportionate to the recipient's time, effort, responsibility, and professional discretion. "You need to mark all of them as exceptional," for example, represents an evaluative judgement as an obligation. Post-editing should acknowledge that the decision remains with the referee rather than merely adding please.

Politeness and mitigation

Internal modification occurs within the main request, while external modification includes supportive moves such as preparators, grounders, acknowledgements of imposition, and expressions of appreciation (Blum-Kulka et al., 1989). L2 writers may rely heavily on a limited set of lexical markers, particularly please, while using fewer syntactic modifiers (Biesenbach-Lucas, 2007). Yet please does not make every directive appropriate, and multiple apologies or elaborate expressions of gratitude may make a simple message unnecessarily long. Politeness must be evaluated through the complete message and its context.

Voice, Ownership, and Authenticity

The fifth dimension requires more attention because its application to AI-assisted institutional communication is less established, although the underlying construct is not new. Work on writing and identity shows that writers construct a discoursal self through accumulated linguistic choices (Ivanič, 1998), while research on academic and second-language writing treats voice as relational, reader-perceived, and genre-situated (Matsuda & Tardy, 2007; Tardy, 2012). Research on stance similarly identifies resources through which writers position themselves toward readers (Hyland, 2005). This chapter therefore does not propose a new concept of voice; its pedagogical claim is that voice becomes especially important when AI supplies many of the linguistic choices through which it is normally constructed. Here, writer's voice refers to the linguistic and interpersonal choices through which a writer presents a version of the self, including vocabulary, formality, sentence style, and the framing of

requests or explanations. In institutional communication, that voice remains relational and context-dependent.

This issue becomes especially important when students transmit AI-generated messages with little revision. Grammatical accuracy and conventional politeness do not necessarily reflect the sender's intended identity or position. In workplace communication, Cardon and Coman (2026) found that medium to high levels of AI assistance could reduce evaluations of sender trustworthiness despite positive impressions of message professionalism. Coman and Cardon (2026) likewise examined professionalism and authenticity as distinct evaluations. These studies concern workplace communication, so transfer to student-faculty interaction requires investigation. They nevertheless motivate attention to how a polished message represents its sender.

Post-editing can partly reduce the distance between writer and message, but its effectiveness depends on the extent and nature of the writer's intervention. In a preregistered study of 81 participants, Baumler et al. (2026) found that post-editing increased the stylistic similarity between LLM-generated drafts and participants' unassisted writing. However, post-edited texts still remained stylistically closer to LLM-generated writing than to participants' independently composed texts and showed reduced stylistic diversity. The study also identified a gap between perceived stylistic authenticity and measured stylistic similarity: participants could regard a text as representative of their style even when detectable LLM traces remained.

Writer's voice is related to, but distinct from, psychological ownership. Voice concerns how the writer is represented in the text, whereas ownership concerns the extent to which the writer experiences the text as their own. Draxler et al. (2024) found that users did not always perceive ownership of AI-generated texts even when they presented themselves as authors, and greater influence over the text was associated with stronger ownership. Gero et al. (2026) also found that ownership varied by stage of AI support: planning assistance produced a smaller reduction than drafting assistance. These findings do not determine the effect of learner-directed pragmatic revision, but they support keeping substantive decisions with the learner and placing contextual planning before generation.

Authenticity here does not mean that a text must be free from AI assistance. Stylistic authenticity concerns resemblance to a writer's habitual linguistic patterns. Pragmatic authenticity is used here as an operational pedagogical term for alignment among intended meaning, projected identity, relationship, and communicative action. Writers' explanations and endorsement offer evidence for examining that alignment, but cannot

establish how recipients will perceive it. Nor does endorsement demonstrate stylistic authenticity: Baumler et al. (2026) found a gap between self-perception and measured style similarity. Learner accounts should therefore be considered alongside the message, its context, and, where feasible, recipient interpretations.

Review of voice can bring the other dimensions together. Learners ask whether the message sounds plausible for the relationship, whether its formality and certainty reflect their intended position, and whether they can take responsibility for its claims and requests. These questions can also prompt a return to planning or diagnosis. The aim is accountable communication, not recovery of a supposedly pure human voice.

Synthesis

Meaning–intention alignment runs through all five dimensions. Automated reformulation may increase or reduce certainty, convert a tentative question into a firm request, or supply details that the writer never provided. Such changes are not merely stylistic. They can alter the speech act and the obligations attributed to either participant. For this reason, pragmatic diagnosis begins with comparison against the writer's actual intention and factual knowledge. A polished sentence should be rejected when the writer cannot verify its content or would not endorse the position it communicates.

The dimensions are analytically distinct but operate together. An acceptable address cannot repair missing information, and mitigation cannot justify prescribing a professional judgement. A complete message may still misrepresent the writer. The framework therefore guides evaluation of how linguistic form, contextual information, interpersonal positioning, and responsibility work together, rather than prescribing features to add mechanically.

The dimensions synthesize work on pragmatic competence (Taguchi & Roever, 2017; Thomas, 1983), requests and email communication (Blum-Kulka et al., 1989; Economidou-Kogetsidis, 2011, 2018), voice (Hyland, 2005; Ivanič, 1998; Matsuda & Tardy, 2007; Tardy, 2012), and AI-assisted voice and ownership (Baumler et al., 2026; Draxler et al., 2024; Gero et al., 2026). They form an instructional synthesis, not a validated taxonomy.

Table 1: *Diagnostic Matrix for Pragmatic Post-Editing*

Dimension	Primary diagnostic question	Typical risk in AI output	Evidence of an informed revision
Purpose and completeness	Can the recipient identify the purpose and act without avoidable clarification?	A fluent message omits the deadline, document, attachment, or requested action.	Necessary and accurate context is added; irrelevant detail is removed.
Recipient and relationship	Does the message fit the participants' roles, history, channel, and local practice?	A generic formal register creates distance, or an informal address assumes familiarity.	Address, opening, and framing are selected for the stated relationship.
Directness and imposition	Does the wording leave proportionate choice and recognize the requested effort?	The draft presents a request as an obligation or becomes so indirect that the action is unclear.	The head act is clear and its force matches time, effort, and professional discretion.
Politeness and mitigation	Do supportive moves help the recipient interpret the request without unnecessary elaboration?	*Please* is added mechanically, or apologies and gratitude are repeated.	Mitigation and support are functional, concise, and distributed appropriately.
Voice, ownership, and authenticity	Can the writer understand, endorse, and take responsibility for the whole message?	Vocabulary, certainty, or formality does not represent the writer's intended position.	The writer accepts, rejects, or reshapes wording through reasoned choice.

Note. Table 1 converts the conceptual dimensions into diagnostic questions and observable revision evidence. It is intended to guide analysis and feedback; it should not be used as a numerical scale or as evidence that one formulation is universally appropriate.

The Pragmatic Post-Editing Cycle

The six-stage cycle operationalizes the framework for situations in which learners translate an initial message or ask AI to generate one. In both routes, the output is treated as a provisional draft. The cycle is a pedagogical proposal rather than a validated assessment model.

Contextual Planning

Before consulting AI, the writer identifies purpose, recipient, relationship, necessary information, channel, and imposition. Making these decisions explicit reduces reliance on the system to infer missing context. Gero et al. (2026) found smaller ownership reductions with planning assistance than drafting assistance in a short-essay study. This indirectly motivates attention to when support is introduced; it does not establish that the proposed planning stage preserves ownership in institutional communication.

AI-Assisted Generation or Translation

The writer may translate a personally composed message or request a draft based on a description of the situation. A useful prompt specifies relevant contextual constraints, but detailed prompting cannot guarantee appropriateness. The resulting text may still be overly formal, unnecessarily elaborate, too direct, or inconsistent with the intended relationship.

Pragmatic Diagnosis

The writer first checks meaning-intention alignment: whether the output preserves intended content without unsupported additions, omissions, or changes in certainty or urgency. The draft is then evaluated through the five dimensions. AI may suggest alternatives, but should not be the sole judge of appropriateness. Andersson and McIntyre (2025) found that ChatGPT 3.5 recognized many instances of impoliteness but had difficulties with some contextual interpretations and register. Their exploratory findings support human review without establishing a uniform limitation across models or tasks.

Informed Revision

Revision addresses a diagnosed problem rather than making the text generally "better" or "more polite." Writers may revise directly or prompt AI with a specific goal, such as retaining WhatsApp brevity while adding information needed by the recipient. Human involvement should not be measured by compulsory word replacement. What matters is whether the writer guides, understands, and evaluates the change.

Final Verification and Authorial Approval

The writer confirms factual details, attachments, names, dates, deadlines, and submission procedures, then checks that the revision has not altered the intended meaning or created a new interpersonal mismatch. Approval means recognizing the message as an acceptable representation of one's meaning and position and accepting responsibility for sending it.

Focused Reflection

A brief reflection identifies the principal problem, the revision made, and why the new formulation better suits the purpose and recipient. The focus is pragmatic reasoning, not a general account of AI use. This stage makes otherwise implicit decisions available for learning and feedback.

Iteration, Prompting, and Process Records

Movement through the cycle is iterative rather than strictly linear. A revision may create a new factual, relational, or stylistic problem, so writers may return to diagnosis or planning. This flexibility prevents the stages from becoming a mechanical sequence while preserving their central purpose: contextual reasoning, informed revision, and accountable approval.

The distinction among diagnosis, revision, and verification remains pedagogically important. Revising too quickly can lead learners to replace visible words without identifying the underlying problem. Naming the problem first makes the reason for change explicit and helps distinguish pragmatic decision-making from general language correction.

Comparison can make that judgement visible. Learners may retain the initial text, first AI output, diagnosis, revised version, and brief explanation, then compare what changed and why. Such process records can support feedback and reflection without becoming compulsory documentation for every real-world message.

Flexibility in Application

The cycle does not require equal time at every stage. Familiar low-stakes exchanges may need only brief planning and verification, while higher-stakes requests may require closer checking. Focused reflection is mainly an instructional scaffold rather than a requirement for every message. The stages are adaptable, but the core questions remain stable: what is intended, what does the recipient need, what changed, and can the writer stand behind the final version?

Table 2: *Observable Learning Evidence Across the Six-Stage Cycle*

Cycle stage	Learner action	Possible record	Teacher's diagnostic prompt
Contextual planning	Defines purpose, recipient, channel, relationship, information, and imposition.	Brief context card	What must be decided before language is generated?
Generation or translation	Uses a self-written source or a context-rich prompt and retains the output.	Prompt/source and first draft	Which decisions did the prompt make explicit?
Pragmatic diagnosis	Identifies a mismatch through the five dimensions.	Annotated draft	What might prevent the recipient from interpreting or acting?
Informed revision	Changes only features connected to the diagnosis; may revise or reprompt.	Revised message	Which change addresses the diagnosed problem?
Verification and approval	Checks facts, attachments, intention, and interpersonal implications.	Final checklist and approved text	Can every claim and requested action be verified?
Focused reflection	Explains the principal problem, change, and contextual reason.	Two- or three-sentence rationale	Why is the revision more suitable for this recipient and purpose?

Note. Table 2 makes learner decision-making visible across the cycle. The records are optional classroom evidence rather than compulsory documentation for every message, and the teacher prompts are designed to elicit contextual reasoning rather than one preferred wording.

Applying the Cycle: Two Composite Tasks

The examples are constructed pedagogical composites, not verbatim messages, systematically collected data, or evidence that an identifiable learner used AI. The intermediate draft in Task One is constructed to illustrate possible AI output; it is not a recorded model response. Email and instant messaging have different conventions, so brevity and informality are not inherently inappropriate.

The examples focus on requests because requests make relationships and imposition especially visible. The Cross-Cultural Speech Act Realization Project distinguished the principal request, or head act, from internal

modification and external supportive moves (Blum-Kulka et al., 1989). This helps explain why adding a polite word may not repair a message: the head act, supporting information, and context must be considered together. The two tasks also show that pragmatic problems differ across levels of imposition and informational need.

Task One: Following Up on a Recommendation Letter

Initial message: "Hi teacher, when can you write my recommendation letter? You told me to remind you." Assume an established WhatsApp exchange and an earlier agreement to provide a reference, but that the deadline and submission route have not yet been shared. The timing question may be reasonable after that agreement. The diagnostic issue is whether the recipient has sufficient information in the conversation, not whether every reminder repeats the full application context.

Contextual planning records the programme, university, deadline, submission route, available materials, prior agreement, and referee's discretion. A possible prompt is: "Draft a short WhatsApp reminder about a recommendation letter already agreed. Include [programme], [university], [deadline], and [submission route], and ask whether the deadline is manageable." This retains the learner's decisions before generation.

Constructed intermediate draft: "Dear Esteemed Professor, I respectfully remind you to complete my recommendation for [programme] at [university] by [date], using [submission route]. Please ensure that all ratings are exceptional. I am profoundly grateful for your invaluable assistance." Diagnosis identifies excessive formality for the stated exchange and an unsupported instruction about ratings. That instruction changes a reminder into pressure on the referee's professional judgement. The learner removes it, selects the locally appropriate address, and checks feasibility rather than prescribing completion.

A possible revision is: "Hi Ms [Surname], I'm following up on the recommendation letter you kindly agreed to write. It is for [programme and university], with a deadline of [date]. The university will send you a submission link. Would that deadline be manageable? I can send my CV and any other information you need. Thank you again." Verification checks the agreement, facts, submission route, and appropriate title; previously shared details may be omitted. The learner then approves the message. A focused reflection might state: "I removed the instruction about ratings because that judgement belongs to the referee. I retained the necessary application details and asked about the deadline." These models reasoned post-editing rather than treating either draft as a universal template.

Task Two: Requesting Assistance with Coursework

Initial message: "Hi teacher, I have homework, but I can't do some parts." This comparatively lower-imposition request could arise in a preparatory or foundation programme. Its main issue is ambiguity rather than obvious impoliteness: the difficult questions and desired help are unspecified. Its actual imposition depends on the work requested and local expectations about instructor availability.

Planning identifies the assignment, questions, prior attempt, attachment, and desired assistance. Diagnosis asks whether that information is already shared or must be supplied. In an AI-assisted version of this task, learners retain the generated draft and check whether it requests explanation, feedback, or answers. The shorter illustration below shows the intended revision outcome; Task One provides the fuller intermediate-draft demonstration.

A possible revision is: "Hi Ms [Surname], I'm having difficulty with questions 6 and 8 in the past-simple homework. I tried both, but I'm not sure which verb forms to use. I've attached a photo of the page. Could you please explain what I should check? Thank you." Final verification confirms the question numbers, grammar topic, attachment, and intended request for explanation. Reflection can note that the revision remains brief but enables a focused response and demonstrates that the learner attempted the task.

Comparing the Two Cases

The two cases show why surface indicators should not be scored in isolation. The first revision is longer because the recipient needs several pieces of application information; the second can remain brief because the task is already shared and the requested help is limited. The presence of please in the coursework message, for instance, matters less than the specification of the questions and the request for explanation — pragmatic post-editing aims for proportionate completeness rather than maximum elaboration or mechanical politeness marking.

Across both tasks, the most visible linguistic change is not necessarily the most important pragmatic change. In the recommendation example, the revised modal question is supported by information that enables the recipient to judge feasibility. In the coursework example, the key improvement is specification of the task and type of help. A relatively high-imposition request benefits from acknowledgement of time and discretion; a lower-imposition request benefits primarily from an explicit, manageable action.

Alternative revisions should remain possible. A learner who regularly addresses an instructor by first name need not adopt a title simply because a model message uses one, and a short messaging exchange may not require an email-style closing. Learners can compare alternatives by holding the situation constant and changing one feature at a time — the form of address, the head act, the amount of supporting information, or the closing — and predicting how each change could affect clarity, distance, urgency, or perceived entitlement.

Pedagogical Applications

The framework can be used as a classroom sequence for university students and adult learners. Its aim is not only to improve one message but to develop the ability to diagnose pragmatic problems and make context-sensitive decisions. Research on instructed second-language pragmatics supports explicit attention to pragmatic meaning, although instruction should not prescribe one universally polite formulation (Jeon & Kaya, 2006).

Lesson Design and a Three-Lesson Unit

A lesson may begin with an anonymized message and a description of the writer, recipient, relationship, channel, purpose, and institutional conditions. Learners first analyze it independently through the five dimensions and then compare interpretations in pairs or small groups. Structured peer interaction can support language awareness when learners are prepared to explain and negotiate their judgements (Sato & Ballinger, 2012). The teacher asks for contextual justification rather than a vote on which version "sounds better."

Learners next generate or translate a message using a prompt that includes relevant contextual information. They compare the initial message, AI-assisted version, and their own diagnosis, noting what the system added, removed, softened, intensified, or made more formal, then revise personally or reprompt with a specific goal. Teacher feedback remains diagnostic at first — for example, indicating that the purpose is clear but information needed by the recipient is still absent. A model is introduced only when necessary and discussed as one contextually defensible option.

A three-lesson unit can move from diagnosis to supported use of the full cycle and then independent transfer. In the third lesson, learners first complete a comparable task with reduced scaffolding, making independence easier to assess. A subsequent task can vary imposition, relationship, or shared information to increase the need for pragmatic reasoning. Changing these demands separately from support helps teachers interpret

performance. This proposed sequence requires classroom evaluation rather than assuming that increasing demands necessarily improves learning.

Recipient-response activities make the communicative outcome observable. One group receives only the message and prepares a feasible reply; another holds the writer's context card. They compare the reply with the intended action and identify missing information or unintended pressure. Repeating the exchange after revision tests whether the message enables the intended response. This information gap connects metapragmatic analysis to meaning-oriented task completion.

Formative Assessment and Differentiation

Assessment is primarily formative and considers both the message and the learner's explanation. The five-dimensional checklist functions as temporary scaffolding rather than a validated rating scale, and prompts should be reduced as learners become more independent. The activity ends with factual verification, authorial approval, and brief reflection. Retaining AI-generated wording is acceptable when it follows conscious evaluation rather than automatic acceptance.

Tasks can be adjusted for proficiency and time. Less experienced learners may compare short versions and answer guided questions, while advanced learners can analyze competing interpretations and defend a final choice. The communication channel should remain visible because an email may require fuller framing than an established messaging exchange. The aim is contextual appropriateness, not converting every message into formal academic prose.

Feedback can be organized around decision quality rather than the number of corrections. Evidence includes whether the learner identified the intended speech act, supplied relevant context, anticipated the recipient's needs, recognized disproportionate force, verified facts, and justified the final choice. Linguistic accuracy remains important because errors may obstruct interpretation, but it is considered alongside pragmatic effect. A simple three-level description — emerging, developing, and independent — may be used for classroom feedback, provided it is not presented as a validated research scale.

At the emerging level, the learner mainly evaluates grammar or accepts AI output without contextual justification. At the developing level, the learner identifies at least one meaningful mismatch and makes a relevant revision but may overlook another dimension or rely heavily on teacher prompts. At the independent level, the learner reconstructs the context, diagnoses interacting problems, selects a proportionate solution, verifies

the message, and explains why it is defensible. These descriptions concern observable performance in a particular task; they should not be treated as fixed characteristics of a learner.

The Teacher's Role and Peer Feedback

Teacher intervention should preserve uncertainty where more than one formulation is defensible. Instead of labelling isolated expressions as universally polite or impolite, teachers can ask what the recipient knows, what action is requested, how much choice the wording permits, and what alternative interpretations are possible. This supports both pragmalinguistic and sociopragmatic awareness while discouraging reliance on a single teacher-approved template.

Feedback can follow the order of the cycle: first missing or unverifiable content, then the relationship and force of the request, and finally linguistic realization. Peer feedback can use the same questions, but judgements should be linked to context. A comment such as 'too direct' becomes more useful when it identifies which wording creates an obligation and why that obligation may be disproportionate in the stated relationship.

Scope, Boundaries, and Extension

The framework can also support cross-linguistic comparison without treating first-language practices as errors. Learners may begin with a message in a language they use confidently and examine whether a literal translation creates the intended effect in the target context. The goal is to expand learners' repertoire of choices rather than replace one cultural norm with another.

Teachers should establish boundaries for tool use before the activity, including permitted uses of AI, anonymization, and disclosure requirements. Where access is unequal, a common AI output can be provided so assessment focuses on diagnosis and revision rather than platform quality. Activities should not depend on paid access or require learners to upload confidential correspondence.

The framework can extend beyond requests, but each speech act requires its own analysis. Apologies involve responsibility and repair; refusals require clarity and relational management; complaints involve evidence and proportionality. Extending the cycle means retaining contextual planning, diagnosis, revision, verification, and accountable approval while adapting the diagnostic questions to the communicative act.

Final Checklist

The following checklist can be used after revision and before authorial approval. It is deliberately phrased as questions rather than rules because

an affirmative answer must be supported by the actual communicative context. In classroom use, learners may mark yes, not yet, or not applicable and briefly explain any item that required a change.

Table 3: *Final Pragmatic Post-Editing Checklist*

Dimension	Review focus	Question for the writer	Evidence to verify
Meaning–intention alignment (cross-cutting)	Intention	Does the message perform the action I actually intend?	The request, enquiry, reminder, submission, or explanation is identifiable.
Meaning–intention alignment (cross-cutting)	Accuracy	Is every name, date, deadline, claim, and procedural detail correct?	Details are checked against the relevant source rather than accepted from AI.
1. Purpose and completeness	Completeness	Can the recipient respond without avoidable clarification?	The necessary task, document, deadline, attachment, and requested action are supplied.
1. Purpose and completeness	Relevance	Have I removed details that do not help the recipient interpret or act?	The message is concise for its purpose and channel without becoming vague.
2. Recipient and relationship	Relationship	Do the address, opening, and overall register fit this recipient?	Choices reflect roles, prior contact, local practice, and the communication channel.
3. Directness and imposition	Imposition	Does the wording recognize the recipient's time, effort, and discretion?	The message does not present optional assistance or professional judgement as an obligation.
3. Directness and imposition	Clarity	Is the central requested or reported action explicit enough?	Indirectness does not force the recipient to infer what response is wanted.

Dimension	Review focus	Question for the writer	Evidence to verify
4. Politeness and mitigation	Mitigation	Do politeness and supportive moves serve a clear function?	*Please*, apologies, explanations, and thanks are proportionate rather than mechanical.
5. Voice, ownership, and authenticity	Voice	Does the language represent how I wish to sound in this relationship?	Vocabulary, formality, certainty, and emotional tone are understandable and acceptable to the writer.
5. Voice, ownership, and authenticity	Responsibility	Can I explain and take responsibility for the complete final message?	The writer has personally reviewed and approved the text and its interpersonal implications.

Note. Table 3 operationalizes the five dimensions as a final-review scaffold; the first column shows which dimension each item serves, and the two meaning–intention items apply across all five. Its value lies in the explanation supporting each response, not in obtaining ten affirmative answers; some items may legitimately be inapplicable in a particular task or channel.

Limitations, Implications, and Conclusion

Limitations

Pragmatic post-editing cannot eliminate uncertainty. Appropriateness remains relational, culturally situated, and open to more than one defensible interpretation. The framework should not establish one variety of institutional English as the universal norm, nor should its dimensions be treated as a standardized scoring instrument. Assessing pragmatic performance requires contextualized judgement that cannot be reduced to isolated linguistic features (Taguchi & Roever, 2017).

AI output may introduce inaccurate information, alter the force of a request, omit qualifications, or reproduce dominant conventions without representing the writer's context. Educational discussions of large language models therefore emphasize critical evaluation, human oversight, bias, and risks of overreliance (Kasneci et al., 2023). Learners must also protect personal information: authentic correspondence may contain

names, identification numbers, application details, and information about others, so unnecessary or confidential data should not be entered into AI systems and anonymized placeholders should be used where possible (Miao & Holmes, 2023).

Access and institutional policy also matter. Learners differ in access to paid systems, connectivity, prompting experience, and teacher support, and may therefore benefit unequally from the same technology (Warschauer et al., 2023). Activities should not depend on a particular commercial platform, and disclosure should follow the requirements of the institution, course, or assessment; pragmatic post-editing is not a method for concealing AI involvement but a means of supporting responsible use.

The cycle is most suitable for teachable communicative tasks in which learners can examine purpose, recipient, and consequences. It should not be presented as a substitute for institutional advice in high-stakes, confidential, legal, or safeguarding communication, nor should every informal classroom message undergo all six stages in full. With practice, some checks may become rapid and internalized; the explicit sequence is scaffolding for judgement, not a requirement to make ordinary communication slow or uniform.

Finally, the proposed framework and cycle have not yet been validated as a unified model across learner groups, languages, genres, or institutions. The examples are pedagogical composites rather than a systematically sampled corpus, and the classroom observations motivating the chapter are not generalizable findings.

Implications for Curriculum and Research

At curriculum level, pragmatic post-editing can connect translation, writing, digital literacy, and language-use courses. Translation tasks can include target-text purpose and recipient response; writing tasks can examine changes in stance and interpersonal force; and digital-literacy work can require factual verification and accountable acceptance. These questions can be embedded in existing activities whenever learners use machine translation or generative AI for communication.

Evaluation should distinguish product quality from learning evidence. A recipient's preference for a revised message does not by itself demonstrate pragmatic development, while a sound diagnosis may still coexist with linguistic difficulty. Studies of the cycle should therefore combine product ratings with process evidence, learner explanations, and recipient interpretations while controlling the communicative situation closely enough to connect changes in appropriateness to identifiable revisions.

Further research should examine independent judgement and transfer across recipients, speech acts, channels, and imposition. Delayed post-tests help distinguish durable development from task-bound performance; Lai et al. (2026) provide an email-pragmatics precedent. Learner diagnoses should also be compared with recipient interpretations. Huynh et al. (2026), using survey data from Vietnamese undergraduates, examined usefulness, trust, professionalism, and attitudes toward AI-assisted institutional emails. This recipient-side evidence motivates separate measures of these constructs; it does not validate the present cycle.

The role of AI should also be varied experimentally rather than treated as a single condition. Planning support, translation, full-draft generation, targeted reformulation, and automated feedback distribute decisions differently between learner and system. Findings on ownership (Draxler et al., 2024; Gero et al., 2026) suggest that these distinctions matter when asking which forms of support improve communication without reducing the learner's ability to reconstruct the context, explain revisions, and communicate independently.

Conclusion

Generative AI does not remove the need for communicative competence; it changes where part of that competence is exercised. When fluent language can be produced quickly, learners must still determine what should be communicated, anticipate how it may be interpreted, verify the content, and decide whether the message represents them. Pragmatic post-editing connects technological assistance with these human responsibilities. Its educational value lies not in guaranteeing a perfect message, but in helping learners make choices they can understand, justify, and accept as their own communicative acts.

Declaration on the Use of Artificial Intelligence

Generative AI tools supported language refinement, paragraph organization, consolidation of overlapping material, bibliographic checking, and revisions to pedagogical illustrations. The intermediate draft labelled as possible AI output is a constructed teaching example, not an empirical model response. The author remains responsible for the conceptual framework, interpretation of sources, pedagogical decisions, and final text. AI assistance does not constitute independent verification of the cited research.

References

Andersson, M., & McIntyre, D. (2025). Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model. *Journal of Pragmatics, 239*, 16–36. https://doi.org/10.1016/j.pragma.2025.02.001

Baumler, C., Bao, C., Nghiem, H., Yang, X., Carpuat, M., & Daumé III, H. (2026). Can you make it sound like you? Post-editing LLM-generated text for personal style. In M. Liakata, V. P. Moreira, J. Zhang, & D. Jurgens (Eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 43867–43895). Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.acl-long.2030

Biesenbach-Lucas, S. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning & Technology, 11*(2), 59–81. https://doi.org/10.64152/10125/44104

Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex.

Bowker, L., & Buitrago Ciro, J. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing. https://doi.org/10.1108/9781787567214

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Cardon, P. W., & Coman, A. W. (2026). Professionalism and trustworthiness in AI-assisted workplace writing: The benefits and drawbacks of writing with AI. *International Journal of Business Communication, 63*(4), 1825–1857. https://doi.org/10.1177/23294884251350599

Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly, 89*(2), 340–359. https://doi.org/10.1177/23294906241233224

Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D., & Welsch, R. (2024). The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors. *ACM Transactions on Computer-Human Interaction, 31*(2), Article 25. https://doi.org/10.1145/3637875

Economidou-Kogetsidis, M. (2011). 'Please answer me as soon as possible': Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics, 43*(13), 3193–3215. https://doi.org/10.1016/j.pragma.2011.06.006

Economidou-Kogetsidis, M. (2018). 'Mr Paul, please inform me accordingly': Address forms, directness and degree of imposition in L2 emails. *Pragmatics, 28*(4), 489–516. https://doi.org/10.1075/prag.17025.eco

Economidou-Kogetsidis, M. (2023). "The language is very formal and appropriate": L2 learners' in/appropriateness evaluations and metapragmatic judgments in student-faculty emails. *Journal of Pragmatics, 217*, 17–32. https://doi.org/10.1016/j.pragma.2023.09.001

Economidou-Kogetsidis, M., Woodfield, H., & Savvidou, C. (2021). Non-native EFL teachers' email production and perceptions of e-(im)politeness. *Journal of Politeness Research, 17*(2), 155–187. https://doi.org/10.1515/pr-2020-0046

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Gero, K. I., Long, T., Schnitzler, C., & Dhillon, P. (2026). From planning to revision: How AI writing support at different stages alters ownership. In *Proceedings of the 2026 ACM Designing Interactive Systems Conference* (pp. 2903–2934). Association for Computing Machinery. https://doi.org/10.1145/3800645.3813003

Huynh, M. K., Tran-Pham, T. A., & Nguyen, H. K. (2026). Student perceptions of AI-assisted institutional emails. *Business and Professional Communication Quarterly*. Advance online publication. https://doi.org/10.1177/23294906261418213

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192. https://doi.org/10.1177/1461445605050365

International Organization for Standardization. (2017). *Translation services—Post-editing of machine translation output—Requirements* (ISO Standard No. 18587:2017). https://www.iso.org/standard/62970.html

Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. John Benjamins. https://doi.org/10.1075/swll.5

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). John Benjamins. https://doi.org/10.1075/lllt.13.10jeo

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Blackwell.

Lai, S., Economidou-Kogetsidis, M., & Savvidou, C. (2026). Long-term effects of pragmatics strategy instruction: Email invitation-requests and Chinese EFL learners. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. https://doi.org/10.1515/iral-2025-0106

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.

Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes, 26*(2), 235–249. https://doi.org/10.1016/j.esp.2006.10.001

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

Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing*. Language Science Press. https://doi.org/10.5281/zenodo.5646896

Nord, C. (2018). *Translating as a purposeful activity: Functionalist approaches explained* (2nd ed.). Routledge. https://doi.org/10.4324/9781351189354

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Sato, M., & Ballinger, S. (2012). Raising language awareness in peer interaction: A cross-context, cross-methodology examination. *Language Awareness, 21*(1–2), 157–179. https://doi.org/10.1080/09658416.2011.639884

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Spencer-Oatey, H. (Ed.). (2008). *Culturally speaking: Culture, communication and politeness theory* (2nd ed.). Continuum.

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.

Tardy, C. M. (2012). Current conceptions of voice. In K. Hyland & C. Sancho Guinda (Eds.), *Stance and voice in written academic genres* (pp. 34–48). Palgrave Macmillan. https://doi.org/10.1057/9781137030825_3

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

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

Part IV
Assessment and Ethics

Assessment in French Language Instruction in the Age of Generative Artificial Intelligence

Gizem Köşker [1]

Abstract
Generative artificial intelligence enables French learners to generate texts, perform translations, receive grammatical corrections, and adapt content to different target audiences. While these capabilities can support learning, they also challenge the validity of assessments based mainly on final products, since a linguistically accurate text does not necessarily reflect a student's independent language performance or understanding. This chapter examines how assessment in French language instruction can be restructured in the age of generative AI, drawing on research on language assessment, action-oriented approaches, authentic assessment, second-language writing feedback, and critical AI literacy. It proposes a four-stage assessment process comprising independent initial performance, documented AI-supported editing, critical validation, and oral explanation with transfer to a new context. Rather than a standardized model, this process is presented as an adaptable design proposal aligned with course objectives and CEFR levels. Sample tasks for A2, B1, and B2, analytical scoring criteria, limits on AI use, and ethical principles are also discussed. The chapter argues for assessment based on multiple forms of evidence that make students' independent language use, decision-making, and communicative responsibility visible.
Keywords: generative artificial intelligence, French as a foreign language instruction, language assessment, process-based assessment, artificial intelligence literacy.

Introduction

In generative AI tools, many tasks that language learners previously performed separately using different applications have been consolidated into a single environment. Thanks to this capability, a student can now write a French text in just a few seconds, have a draft corrected, ask about differences in word usage, adapt a text to a different level, or engage in various activities with a virtual conversation partner. The fact that these tools generate responses that closely resemble natural language offers new possibilities in foreign language instruction in terms of feedback, practice, and material development (Chapelle, 2025; Kohnke et al., 2023). These same features also require a more careful and detailed examination of the

[1] Asst. Prof. Dr., Anadolu University, Faculty of Education, Department of Foreign Language Education, French Language Teaching Program, ORCID: 0000-0002-0115-9756, *gizemkosker@anadolu.edu.tr*, Türkiye

connection between a student's submitted work and their independent language proficiency.

At this point, the problem encountered in language assessment is not merely that AI can write text in the student's place. Generative AI can also enhance the content of a text, completely alter its language level, correct errors, make the discourse more fluent, and mimic certain language usage patterns. Therefore, the apparent success of a written text, translation, or presentation prepared at home may not indicate that the student can use that language independently. Current research on language assessment indicates that generative AI can be beneficial in essay writing, grading, and providing feedback, while also highlighting that it can raise new issues regarding validity, fairness, transparency, and human oversight (Chuang & Yan, 2025; Jang & Sawaki, 2025).

This shift is also prominent in French language instruction. Proficiency in French cannot be measured only by the correct application of grammar rules (Yılmaz Güngör, 2020). Students are expected to construct a text appropriate to their communicative purpose, choose the correct form of address, adjust the level of formality, respond to the addressee's reaction, and explain cultural content in an understandable manner. In this context, the Companion Volume to the Common European Framework of Reference for Languages recognizes the student as a social actor who uses language in real-life tasks; it places dimensions such as interaction, mediation, and adaptation to the target audience at the center of language proficiency (Council of Europe, 2020). An output corrected by artificial intelligence can make these dimensions visible. However, this alone is not sufficient to prove that the student can perform all of these tasks independently.

The aim of this section is not to advocate for either banning or unrestricted use of generative AI in French language instruction. The goal is to determine the principles on which assessment criteria should be based when AI is used. To this end, the opportunities and risks that generative AI brings to language assessment will first be addressed, followed by a discussion of the limitations of product-oriented tasks and AI detection tools. Subsequently, the theoretical foundations of assessment design will be explained, and a four-stage implementation process for French language instruction will be presented. The final sections will focus on tasks for the A2, B1, and B2 levels, an analytical scoring approach, the teacher's role, and ethical principles. The proposed process is not a new, empirically validated assessment model. Rather, it is an application design for French language instruction based on the existing literature on assessment and AI literacy.

Why Is Generative AI Changing Language Assessment?

New Possibilities and New Validity Issues

Unlike traditional automatic correction tools, generative AI does not simply flag specific errors. It can generate a new text from scratch based on the user's intent, take into account previous stages of the conversation, and reorganize the same content in different ways. From a language education perspective, this supports functions such as personalized practice, real-time feedback, role-playing, vocabulary and grammar explanations, text simplification, and material preparation (Chapelle, 2025; Godwin-Jones, 2024; Kohnke et al., 2023). Thanks to this, students can independently ask questions tailored to their own level, request further clarification on points they do not understand, and attempt written or oral exercises without fear of making mistakes.

However, when these same features are used for assessment purposes, a more complex picture emerges. If the task assigned to a student can be largely completed by a tool on the student's behalf, the meaning of the score received becomes entirely unclear. A text submitted by a B1-level student that is at the B2 or C1 level may reflect not the student's language development, but rather the capabilities of the tool used at that stage. Even if text generated by artificial intelligence appears correct, it may still contain issues regarding content, source, linguistic register, or communicative context. Chuang and Yan's (2025) systematic review emphasizes that while generative artificial intelligence enhances efficiency and feedback opportunities in language assessment, it simultaneously raises concerns regarding reliability, validity, fairness, and feasibility. Jang and Sawaki (2025) also underscore the need to maintain human oversight and interpretability in both automated scoring and interactive assessment.

At this point, it is necessary to distinguish between two different uses of the assessment tool. In the first use, AI assists the teacher in preparing items or materials and generating feedback. The key question here is the accuracy of the content generated by the tool, as well as its alignment and consistency with the structure being measured. In the second use case, the student utilizes artificial intelligence to complete an assessment task. Here, the focus shifts to what the student contributes to the product, the decisions they make, and how their independent competence is demonstrated. These two use cases do not give rise to the same ethical and assessment issues. Imposing a blanket ban or granting blanket permission regarding "the use of AI" in every task obscures this distinction.

When conducting assessments in French language instruction, the feedback provided by AI must also be handled with care. While large language

models can often generate fluent French, this does not always guarantee linguistic fluency or pragmatic and cultural appropriateness in every context. Godwin-Jones (2024) states that generative AI can be a powerful aid in language learning, but its processing of the social and cultural nuances of human language through statistical patterns creates limitations. For this reason, teachers should treat AI output not as a ready-made correct answer, but as a suggestion that needs to be verified and evaluated according to context.

The Limitations of Product-Focused Tasks

For a long time, foreign language classes have relied on texts written at home, translation assignments, and presentations. These tasks allow students to conduct research, draft their work, and revise their texts. However, grading only the final text does not reveal how the product was created. Research on writing in a second language shows that the impact of feedback is related to how it contributes to the student's revision process, and that examining only the revised product is insufficient for explaining the learning process (Hyland & Hyland, 2006). The portfolio approach, in line with these views, advocates for the joint evaluation of the first draft, feedback, rewriting, and reflection (Hamp-Lyons & Condon, 2000). In brief, generative AI has brought this debate into sharper focus.

For example, when a student is asked to write a 250-word opinion piece on online shopping, the tool does more than just generate the text; it can shape the text according to detailed prompts such as B1 level, academic tone, specific conjunctions, or the number of errors. When a teacher sees only the final text, they cannot know whether the student generated the ideas themselves, whether they understood the text, or whether they could defend the same viewpoint orally. Similar issues arise in translation assignments. The final translation may be correct, but it remains unclear how the student decided among alternatives, how they chose their words based on context, and to what extent they understood the meaning of the source text.

Therefore, the issue is not about abandoning certain types of assignments entirely. Writing, translation, and presentation assignments at home retain their value for learning. What needs to change is the practice of considering these assignments alone as proof of competence. Assessment should include not only the final product but also the student's independent work, drafts, justifications for revisions, and performance demonstrated during interactions. The assessment reform report prepared for TEQSA recommends creating diverse assessment evidence at the program level in the age of generative AI, making the learning process visible, and collecting reliable performance samples at meaningful points (Lodge et al., 2023).

Why are bans and AI detection tools insufficient?

Imposing a complete ban on AI use may be necessary in some in-class assignments to observe students' independent performance. However, applying a blanket ban across all assignments is inadequate for two reasons. First, students have easy access to these tools outside of class. Second, using AI critically and ethically is a competency that is becoming increasingly important in higher education and the workplace. Developed as an Assessment Design Scale, the AIAS (Artificial Intelligence Assessment Scale) proposes clearly defining different levels of AI use based on learning outcomes, rather than reducing AI use to a binary choice of "prohibited" or "permitted" (Perkins et al., 2024). This approach helps determine in advance where human input is required for each assignment and how the tool can be utilized.

AI detection tools used to enforce bans also fail to provide a sufficiently reliable solution at this point. Weber-Wulff et al. (2023) demonstrated that widely used detection tools cannot consistently distinguish text generated by AI, and that simple rewriting processes significantly alter the results. Liang et al. (2023), on the other hand, revealed that detection tools can incorrectly classify texts by authors whose native language is not English as AI-generated. Although the studies were conducted in English, the fundamental warning for foreign language learners is clear: texts with low linguistic diversity or those that are more predictable may be incorrectly flagged as suspicious. For this reason, it is incorrect to use a detection score as standalone evidence of academic misconduct.

Focusing on design rather than detection is a more robust and reliable approach. Students may be asked to prepare their first draft in class, explain the changes they made, cite the sources they used, and rephrase the text during a brief oral discussion. These methods do not definitively identify AI use, nor is that the goal. The goal is to gather supporting evidence so that the grade can be linked to the student's learning outcomes.

Foundations of Assessment Design

Validity, Authenticity, and Multiple Lines of Evidence

The most fundamental question regarding an assessment task is which of the student's skills is being measured. If the learning outcome is "the ability to write a text expressing an opinion at the B1 level," it is not sufficient to look only at the final text that has been corrected from a grammatical standpoint. The student must form an opinion, organize the reasoning behind it, use conjunctions, tailor the text to the target audience, and explain

these decisions. When artificial intelligence takes over some of these processes, the assessment may fail to measure the structure originally intended. This is precisely where the issue of validity in language assessment arises (Chuang & Yan, 2025; Jang & Sawaki, 2025).

Authentic assessment requires students to complete a task that is meaningful and connected to real life. However, similarity to real life alone is not sufficient. The task must ensure that the student applies the targeted knowledge and skills, the assessment criteria must be completely transparent, and performance must be interpreted only within the context. In this context, Sotiriadou et al. (2020) have shown that phased and interactive authentic tasks can support skill development and academic integrity. In the age of generative artificial intelligence, authenticity does not mean ignoring the tool. If artificial intelligence can be used in real life, students can be given the opportunity to use it in a controlled manner; however, human decisions that cannot be delegated to an AI tool must also be evaluated.

The multiple evidence approach strikes this balance. The student's independent in-class performance, the product developed at home, the record of AI interaction, the reflective explanation, and the oral response each demonstrate different dimensions of the same skill. None of these pieces of evidence is flawless on its own. In-class writing may be affected by time pressure; oral performance may be influenced by fear of making mistakes; and the process record may be incomplete. Nevertheless, evaluating the evidence collectively provides a stronger foundation than relying only on a single final product. Lodge et al. (2023) recommend collecting evidence at various meaningful points throughout the program rather than placing the burden of assessment validity on a single exam.

Action-Oriented Approach and Mediation

The Companion Volume of the Common European Framework of Reference for Languages provides an important foundation for the design of French language assessment. The framework defines the learner not simply as a person with language knowledge, but as a social actor acting with a specific purpose. Production, reception, interaction, and mediation are not viewed as isolated skills, but rather as communicative activities that can be used together within real-world tasks (Council of Europe, 2020). This approach also clarifies why a single written product is insufficient in the age of artificial intelligence. In addition to creating a text, students may be expected to answer a recipient's question, rephrase the content, or adapt it for an entirely different target audience.

Mediation is particularly well-suited for AI-supported assessment at this point. Students may be asked to explain a French news article to a friend with a lower language proficiency level, convey basic information about a Turkish announcement to French-speaking students, or introduce a cultural practice in context. AI can provide initial suggestions but the student must verify the accuracy of the information, the adequacy of the explanation, and its suitability for the target audience. Over time, these tasks shift the focus from finding word-for-word translations to developing the ability to reconstruct meaning.

The Common European Framework of Reference for Languages (CEFR) levels require that expectations for evaluating AI output be scaled accordingly. An A2 student can identify and explain two changes in a simple invitation message. A B1 student can choose between suggestions regarding text structure and conjunctions. A B2 student may be expected to examine more complex issues such as linguistic register, target audience, source reliability, and cultural generalization. Thus, AI literacy is not evaluated as an additional task independent of the student's language level, but rather within the framework of communicative actions appropriate to that level.

AI Literacy and Process Transparency

AI literacy is not simply about being able to write an effective command to the tool. Ng et al. (2021) define AI literacy in terms of knowing and understanding the system, using it, evaluating it, and considering ethical issues. From the language learner's perspective, all these skills require questioning word choice, linguistic register, information accuracy, and communicative intent, rather than simply accepting the AI's suggestion as correct. A student may use a tool, but the responsibility for the final decision always lies with them.

Wang and Wang (2025), who focus on AI-supported writing in a second language, demonstrate that critical AI literacy is related to awareness, how the student positions themselves in relation to AI, the strategies they use, and how they evaluate the output. This finding shows that asking only "What prompt did you type?" is quite insufficient for evaluation. It is also necessary to simultaneously examine why the student used a particular AI function, how they tested the suggestions, and how they maintained authorship responsibility over the text.

Process transparency serves this purpose. It does not make sense to require students to submit their entire chat history. Instead, the first draft, two or three key prompts used, accepted suggestions, a rejected suggestion, and a brief rationale may suffice. These records are used not to monitor the

student but to explain the creation of the final product. The AIAS also advocates for clearly communicating the limits of AI use in assignments and ensuring that students know in advance what is acceptable (Perkins et al., 2024).

Program-Level Consistency

In the age of generative AI, assessment integrity cannot be confined to a single assignment. Just as tying a course's entire grade to a text written at home does not capture the full picture, transforming the entire assessment into a proctored exam also reflects only a portion of learning outcomes. Lodge et al. (2023) recommend that assessment reform be addressed at the program level rather than through individual assignments. A deliberate balance must be established throughout the student's development between independent, supported, collaborative, and AI-assisted work. While some assignments demonstrate what a student can do without external support, others can assess skills such as using resources, leveraging feedback, and working with technology.

In the French curriculum, this distribution can be planned from the very beginning of the term. Initially, short independent writing and simple speaking samples are collected. During the term, development tasks are assigned that include drafts, peer or teacher feedback, and limited AI support. At a more advanced stage, students critique AI output, verify it against sources, and adapt it for different target audiences. At the end of the term, a single assignment that combines written and oral performance may be required. A systematic process that progresses in this way significantly reduces the pressure to both support learning and ensure high safety standards within a single assignment.

This program-level planning also helps students better understand the rules. When one teacher in the same program completely bans AI while another allows its unrestricted use, it is unlikely that students will understand the boundaries of academic integrity. At this point, AIAS recommends that the permitted levels of AI use in assignments be specified in a common and understandable language (Perkins et al., 2024). The institution or program can prepare common guidelines for types of use that are prohibited, limited, or open-critical. Instructors make their own decisions based on the course's learning outcomes; the common language prevents students from encountering different terminology in every course.

This approach does not require the complete elimination of independent exams. If important decisions regarding written and oral proficiency are to be made, secure samples must be collected under supervised and accessible conditions. However, these samples should never replace skills such

as extended writing, mediation, collaboration, and the critical use of technology. A valid French assessment program must demonstrate what students can do both independently and by making responsible use of contemporary tools.

A Four-Stage Assessment Process for French

The process presented below is not an independent model with this name that has been validated in the literature. It is a process developed by integrating the action-oriented approach of the Common European Framework of Reference for Languages, process-based writing, authentic assessment, AI literacy, and proposals for assessment reform in the age of AI into French language instruction (Council of Europe, 2020; Hamp-Lyons & Condon, 2000; Lodge et al., 2023; Ng et al., 2021; Sotiriadou et al., 2020). The process consists of four stages: independent initial performance, documented AI-supported revision, critical validation, and oral explanation and transfer to a new context. Teachers are not required to apply all four stages with equal weight in every task.

Initial Independent Performance

In the first stage, the student demonstrates the targeted language skill without AI support. This performance can be assessed in the classroom, within a specific time frame, and under pre-defined resource conditions. The task does not need to be very long. The student can write a short email, describe a simple image, respond to an opinion, or verbally explain the main idea of a text. The goal is not to create a punitive testing environment, but to obtain a reliable initial sample of the student's independent language use.

The independent performance should be kept separate from the final product. The number of errors in the first draft should not, on its own, determine the grade. This draft demonstrates the student's vocabulary, grammatical control, text structure, and communication strategies. When compared to subsequent revisions, it provides insight into the learning process. Lodge et al. (2023) recommend collecting examples of confident performance at key points for assessing student progress and proficiency. In a French class, this could take the form of short in-class writing activities or interactive speaking activities.

The task conditions must be clear. Whether or not the use of a dictionary is permitted, the time limit, expected length, and assessment criteria must be clearly communicated in advance. The teacher should not view this stage in which artificial intelligence is not used, as "real performance," while regarding subsequent stages as simple supplementary exercises.

Each stage demonstrates a different learning outcome. The first stage demonstrates independent proficiency, while the subsequent stages demonstrate the ability to utilize feedback, evaluate, and apply what has been learned.

Documented AI-assisted revision

In the second stage, the student refines their initial draft with the help of AI. The scope of AI use is determined based on the assignment's objective. The teacher may allow AI assistance only for grammar and spelling checks; may request feedback on text structure, word choice, or target audience; or may assign the task of comparing different AI responses. Perkins et al. (2024) state that clarifying these usage levels before the task begins establishes a common language between the teacher and the student.

At this stage, students should be encouraged to ask the AI specific questions rather than having it completely rewrite the text. For example, in a B1 writing task, commands such as "Rewrite the text; highlight only issues related to the use of conjunctions" or "Explain three grammar errors while maintaining the B1 level" can be used. This limitation keeps the cognitive load on the student. AI serves as a source of feedback but does not become the author of the text. The literature on writing in a second language emphasizes that for feedback to translate into learning, the student must be involved in the decision-making process regarding corrections (Hyland & Hyland, 2006).

The process log should be brief and functional. The student can present the name of the tool used, its purpose, the most important commands, and the key changes made to the text in a one-page form. At this stage, the length of the log should not increase the assessment score. What is being evaluated is the clarity of the process. The student's decision not to use AI is not automatically grounds for a low score. If the use of artificial intelligence is part of the assignment's learning outcomes, an equivalent alternative must be provided; if it is only optional, the student must be able to demonstrate the same skills without using the tool.

Critical evaluation

In the third stage, the student evaluates the AI's suggestions rather than applying them directly. The evaluation should not be limited to linguistic accuracy. Questions such as whether the suggested phrasing is appropriate for the student's level, whether it alters the meaning of the text, whether the tone is consistent with the intended audience, whether the information is verifiable, and whether the cultural explanation is overly general can be addressed. The studies by Ng et al. (2021) and Wang and Wang (2025)

demonstrate that evaluation and ethical inquiry are fundamental components of AI literacy.

In French instruction, linguistic register is a key area for verification. Indeed, AI may suggest a grammatically correct expression that is, however, too formal, too informal, or beyond the student's level relative to the target audience of the task. A student might reject a more formal academic phrase offered in place of "Je pense que" because it does not align with the overall tone of a B1-level text. Similarly, in an invitation message, a student might choose not to use a formal form of address because the recipient is a close friend. The value of these decisions lies not in whether they are accepted or rejected, but in the reasoning behind them.

Validation should be applied more broadly in cultural contexts. Generative AI may present social conventions as uniform rules or use definitive statements without explaining the context. UNESCO lists the protection of cultural and linguistic diversity, human agency, and pluralistic expressions among the fundamental principles for the use of generative AI in education (Miao & Holmes, 2023). In a B2-level mediation task, the student must verify a cultural claim using a reliable institutional or academic source and limit the use of generalizing statements. Asking another AI about an AI's response should not be considered independent verification.

Oral Explanation and Application to a New Situation

In the final stage, the student explains the work they have prepared during a brief interaction. The purpose of the oral discussion is not to detect AI use or to blame the student. The goal is to assess whether the student understands the text, can explain their language choices, and can use French in a similar communicative context. When well-designed, oral assessments can provide valuable evidence regarding validity and academic integrity; however, they require a standardized question structure, clear criteria, and preparation to reduce student anxiety (Nallaya et al., 2024).

The interview may last three to five minutes. The teacher may ask why the student chose a particular phrase from the text, which AI suggestion they rejected, or how they would write the same message to a different person. Selecting some questions from a common pool can enhance scoring consistency. The student should not be expected to recite the text from memory. Asking the student to rephrase a statement in simpler terms, respond to a counterargument, or adjust the language style when the target audience changes produce more meaningful evidence.

Transferring the text to a new context allows the student to use the language without relying only on the prepared text. A complaint email can be transformed into a phone conversation; a cultural explanation can be

simplified for a child; a personal message can be converted into a formal business email. The action-oriented and mediation-based approach outlined in the Common European Framework of Reference for Languages treats the adaptation of linguistic resources to different situations as a key component of proficiency (Council of Europe, 2020). Thus, the oral phase not only serves as a means of identity verification but also assesses an independent dimension of communicative proficiency.

Integrating the Stages into a Single Task

The four stages should not be reduced to isolated, disconnected mini-tests. The student prepares a first draft, receives feedback on the same piece, evaluates the changes, and explains the final product orally. In this way, each stage builds upon the previous one. The teacher, too, tracks the development of the same task rather than comparing different products. This is a form of portfolio assessment. The fundamental strength of portfolio assessment lies not simply in collecting products, but in making visible the changes between drafts and the student's thinking regarding these changes (Hamp-Lyons & Condon, 2000).

When designing a task, the core action to be assessed must first be identified. "Writing a text" is not clear enough on its own. Will the student be persuading a friend, making a formal request, conveying information to someone else, or comparing two viewpoints? Once the communication goal is clear, it becomes easier to determine in which areas AI can be used. Generating ideas, editing language, finding sources, and writing the final text are not the same types of assistance. Teachers should leave tasks directly tied to learning outcomes to the students and should not restrict in the assignment instructions the tasks for which the tool can provide support.

The second decision concerns the evidence to be collected at each stage. Requiring too much documentation can cause students to stray from the core language task and leave teachers facing a workload they cannot manage. For most assignments, the first draft, the final product, three selected AI interactions, and a brief oral discussion are sufficient. Instead of requiring the student to record the entire screen history, the student can be given the responsibility of selecting the examples that best illustrate the decision-making process. This selection also provides insight into metacognitive awareness.

The third decision concerns how to interpret the progress achieved with AI support. A significant difference between the first draft and the final product does not always indicate positive learning. If the student cannot explain the changes or cannot use similar language in a new context, the

progress may be largely attributable to the tool. Conversely, making only a few changes to the final text is not necessarily an indication of failure. The student may be able to justify that the initial draft was appropriate for the purpose. Therefore, it is the appropriateness of the student's choices but not the amount of revision that should be evaluated (Wang & Wang, 2025).

The final decision concerns whether the oral component is task-related. Asking the student general questions unrelated to the text provides limited insight into their ownership of the writing performance. Questions should be linked to the ideas in the text, the linguistic register used, rejected proposals, and potential changes in the target audience. When oral assessment is well-structured, it ceases to be a mere repetition of the written product and instead facilitates a new communicative performance (Nallaya et al., 2024; Sotiriadou et al., 2020).

Table 1. *The four-stage assessment process and its theoretical foundations*

Stage	**What is expected of the student**	**Evidence collected**	**Main theoretical basis**
Independent initial performance	Completing a short task without artificial intelligence support	An example of writing, speaking, or mediation in the classroom	Council of Europe (2020); Lodge et al. (2023)
Documented AI-Supported Revision	Using the tool within specified limits and demonstrating key changes	First draft, selected prompts, final product	Hamp-Lyons and Condon (2000); Hyland and Hyland (2006); Perkins et al. (2024)
Critical validation	Reviewing proposals in terms of language, context, knowledge, and ethics	Acceptance/rejection decisions and brief justifications	Miao and Holmes (2023); Ng et al. (2021); Wang and Wang (2025)
Oral Explanation and Retelling	Explaining the product, answering follow-up questions, and adapting to a new context	Short oral discussion and retelling task	Council of Europe (2020); Nallaya et al. (2024); Sotiriadou et al. (2020)

Examples of Application According to the Common European Framework of Reference for Languages Levels

The examples below demonstrate how the same assessment logic can be simplified across the different levels of the Common European Framework of Reference for Languages. These tasks are not ready-made test items, but examples that teacher can adapt to their own curriculum. The expectations are graded in accordance with the Common European Framework of Reference for Languages' descriptors for production, interaction, and mediation (Council of Europe, 2020).

Level A2: Daily Invitation Message

Students are asked to write a 60–80-word message inviting a French-speaking friend to an event this weekend. The instructions might be as follows: "Vous souhaitez inviter un(e) ami(e) francophone à une activité ce week-end. Écrivez un message de 60 à 80 mots. Présentez l'activité, indiquez le lieu et l'heure, puis demandez à votre ami(e) s'il/elle peut venir." The first draft is prepared in class without the use of artificial intelligence. The basic criteria include providing information about the activity, location, and time; using an appropriate greeting and closing; ensuring clarity; and using simple sentences.

In the second stage, the student receives only limited feedback from the AI. Commands such as "Corrige seulement les erreurs de grammaire. Ne change pas le niveau A2" or "Propose deux formules simples pour terminer le message" can be used. The student selects two changes and explains them briefly. The explanation can be in Turkish or simple French. Detailed cultural analysis is not expected at this level. It is sufficient for the student to notice the change and describe its basic impact on the meaning.

In the final stage, the teacher asks where and at what time the activity takes place. The teacher then assigns a small communication task, such as a time change: "L'activité commence maintenant à 18 heures. Informez votre ami de ce changement." The student is expected to produce a new, short, and clear expression rather than simply repeating the memorized message. Scoring at the A2 level should be based more on the achievement of the communicative purpose than on grammatical perfection (Council of Europe, 2020).

Level B1: Opinion Essay

At the B1 level, students may be asked to write a 150–180-word opinion essay on the use of cell phones in university classes. The text should include the opinion, at least two supporting reasons, an example, and a brief

conclusion. The first draft is written in class. The teacher focuses on the clarity of the opinion, the relevance of the reasons to the topic, the use of basic conjunctions, text organization, and overall comprehensibility. Errors are not evaluated with equal weight as long as they do not completely hinder communication.

In the AI-supported phase, instead of having the student completely rewrite the text, specific feedback is requested : “Indique les passages qui ne sont pas clairs, mais ne réécris pas le texte”, “Vérifie l’emploi des connecteurs logiques” or “Explique les trois erreurs grammaticales les plus importantes.” The student reviews the three suggestions. They may accept the addition of a conjunction; they may reject an expression that is above the B1 level and does not fit the tone of the text. In a brief process note, they explain which suggestions they used and what the most significant change was. This approach treats feedback not as a ready-made correction but as a rewriting process in which the student makes decisions (Hyland & Hyland, 2006; Wang & Wang, 2025).

During the oral discussion, the student explains their strongest rationale, states why they did not use a suggestion they rejected, and responds to the counterargument presented by the teacher. For example, the question “Certains enseignants pensent que les téléphones empêchent totalement les étudiants de se concentrer. Que pouvez-vous répondre?” could be used. This stage also reveals whether the student has taken ownership of the text and can sustain their argument in an impromptu interaction.

Level B2: Intercultural Mediation

At the B2 level, the student prepares a 250–300-word text explaining a cultural practice in Turkey to French-speaking international students. The task requires not only conveying information but also taking the target audience into account and avoiding cultural generalizations. The instructions might read : “Présente une pratique culturelle de Türkiye à des étudiants francophones. Explique son contexte et sa signification, puis indique ce qu’un visiteur devrait savoir. Évite les généralisations et adapte le texte à un public qui ne connaît pas cette pratique.” This task relates to the “mediation of texts and concepts” dimension of the Common European Framework of Reference for Languages (Council of Europe, 2020).

After the first draft, the student asks the AI for feedback on the target audience and clarity. The student does not ask the tool to rewrite the text but rather to highlight points requiring clarification and potential generalizations. The student reviews at least three suggestions. The student verifies a cultural claim using an independent source, qualify definitive and generalizing statements, and adjust the language to suit the target

audience. At this stage, the fact that the information provided by the AI appears fluent does not constitute proof of accuracy. UNESCO's human-centered approach requires the preservation of cultural and linguistic diversity, as well as the reliability of data and sources (Miao & Holmes, 2023).

During the oral presentation, the student explains why they chose the specific application, which generalization they corrected, and how they verified the information. In the retelling task, the student rephrases the same content for a 10-year-old child, a formal academic audience, or a 30-second tourist video. What is being assessed here is not merely vocabulary richness. The student is expected to be able to adjust the level of explanation, tone, and chosen details while preserving the meaning.

Table 2. *Grading of expected performance at the A2, B1, and B2 levels*

Level	Independent task	Use of artificial intelligence	Critical analysis	Oral presentation
A2	Short daily messages	Simple feedback on errors and expressions	Noticing two basic changes	Answering short questions and updating information
B1	Reasoned opinion text	Feedback on coherence, conjunctions, and grammar	Accepting or rejecting suggestions and providing a brief justification	Defending one's opinion and responding to opposing views
B2	Cross-cultural mediation appropriate for the target audience	Feedback on content, format, and target audience	Verification based on linguistic, cultural, and source-related factors	Adapting content to different target audiences and communication contexts

Grading and Assessment Principles

Analytical Grading

Placing too much emphasis on the linguistic accuracy of the final product in AI-assisted tasks carries the risk of mistaking the tool's powerful correction features for the student's actual proficiency. Therefore, independent performance, the final product, process transparency, critical verification, and oral presentation should be evaluated as separate dimensions. An analytical rubric is useful in showing where a student is strong and where

they need support. Criteria should be shared before the task begins, and clear, descriptive definitions should be used for each level.

The following scoring distribution is provided as an example: independent initial performance accounts for 25 percent, the final product for 20 percent, process transparency for 10 percent, critical validation and justification for 20 percent, and oral explanation and presentation for 25 percent. In a course with a strong emphasis on speaking, the oral component may be increased, while in translation or mediation assignments, the verification component may be emphasized. Grading should be based on the quality of the decision, not on whether the student accepts or rejects the AI's suggestion. Consciously accepting a suggestion can also constitute a critical decision.

***Table 3**. Sample Analytical Rubric for Generative AI-Supported French Tasks*

Dimension	**4- Strong**	**3- Adequate**	**2- Developing**	**1- Limited**
Independent Performance	Completes the task in a manner appropriate to the level, clearly and largely independently.	Fulfills the basic task; errors do not significantly hinder communication.	The task is partially completed; meaning becomes unclear at times.	The task's primary objective is not met, or performance is very limited.
Final Product	The product is coherent, appropriate for the target audience, and has been meaningfully developed from the initial draft.	The product is generally understandable and appropriate for the task; some issues persist.	There is limited development; coherence or consistency issues are evident.	The product is inconsistent with the student's level and the task's purpose, or shows no improvement.
Process transparency	The tool, purpose, selected commands, and major changes are clearly shown.	The process is largely understandable; there are minor omissions.	The scope of use is only partially understandable.	The AI's contribution and the student's contribution cannot be distinguished.
Validation and Justification	Evaluates suggestions in terms of language, context, target audience,	Examines most suggestions using appropriate criteria and	The review is superficial; it focuses mostly on	Uses suggestions without questioning them or fails

Dimension	4- Strong	3- Adequate	2- Developing	1- Limited
	and, if necessary, sources; explains their decisions.	provides a basic rationale.	correcting errors.	to provide a rationale.
Oral Explanation and Presentation	Explains the product, answers follow-up questions, and presents the content in a way appropriate to the new context.	Explains the product and key decisions; presents the content with partial support.	Explanation is limited; struggles with follow-up questions and presentation.	Cannot explain the product or use meaningful language in the new context.

To ensure reliable use of the rubric, joint scoring should be conducted on sample student work. If there are multiple teachers, the interpretation of the criteria can be compared. For oral assessments, question pools of similar difficulty should be used; student anxiety, the characteristics of second-language learners, and accessibility needs must be taken into account. Nallaya et al. (2024) demonstrate that the validity and reliability of oral assessments depend on design, structure, clear expectations, and adequate preparation.

Clarifying the Limits of AI Use

The limits of AI use must be clearly specified for each task. Three basic levels may be sufficient. At the first level, AI is not used; the goal is to assess independent performance. At the second level, only specific functions such as grammar checking, word suggestions, or feedback, are permitted. At the third level, AI is now explicitly part of the assignment. The student compares, verifies, and revises the output. This classification is based on the AIAS principle of defining different levels of use according to learning outcomes (Perkins et al., 2024).

The assignment instructions should not simply state, “AI may be used.” It must be clearly specified which tools may be used, at which stage, and for what purposes; what the student is expected to submit; and how the use of AI will be graded. The rules should not be changed afterward. At this stage, the student may be asked to provide a brief statement of use. This statement explains how they prepared the first draft, in which areas they received assistance from the tool, which suggestions they did not use, and how they assumed final responsibility. Punishing a student who demonstrates transparency encourages them to conceal their use of AI.

The Teacher's Role

The teacher's role is to be an assessment designer rather than an inspector trying to catch students who use AI. The teacher defines the learning outcomes, selects the stages where human input is required, prepares age-appropriate tasks, and teaches students how to verify AI output. Language teachers need not only technical knowledge of commands but also AI literacy to make decisions regarding linguistic register, pragmatic appropriateness, source reliability, and ethical use (Kohnke et al., 2023; Miao & Holmes, 2023).

Teachers should not allow AI suggestions to replace their own expertise. Any response that raises doubts regarding French word usage, forms of address, regional variations, or cultural content should be cross-checked against reliable dictionaries, corpora, and institutional sources. Students should also be encouraged to develop the same verification habits. AI feedback is not an automated and impartial equivalent of teacher feedback but rather a supplementary resource selected according to the learning objective.

Ethics, Equity, and Data Security

The protection of personal data is a fundamental requirement in AI-assisted assessment. Students should not upload their own or others' identity, contact, health, or educational information to open systems. Teachers should also not transfer student assignments to third-party systems without explicit consent and institutional approval. Miao and Holmes (2023) note that educational institutions must develop policies that protect data privacy, individual autonomy, equality, and cultural and linguistic diversity. Institutions should prioritize the use of approved tools and provide alternatives for students who do not wish to use personal accounts.

Differences in access can sometimes affect grading. Paid versions may offer more advanced features; some students may have faster internet, better devices, or more experience writing commands. Limiting the weight of the final product in the grade, providing common tools and sample commands, offering basic usage training, and providing equivalent assignments for students who do not wish to use AI can help reduce this inequality. The ability to use AI should not replace French proficiency unless it is explicitly defined as a learning outcome.

Teacher workload should also not be overlooked. Reviewing all chat logs and conducting lengthy interviews with every student for each assignment is neither realistic nor sustainable. All four stages can be applied to one or two core assignments over the course of the semester. The process form can be limited to one-page, oral interviews can be kept to three to five

minutes, and these can be spread out over several weeks. Nallaya et al. (2024) highlight the benefits of oral assessments while also drawing attention to issues of standardization, anxiety, and workload. Therefore, implementation must be aligned with class size and class duration.

Feedback, Self-Assessment, and Student Participation

An analytical rubric should not be used only as a tool for the teacher to assign grades. Students can review the criteria before submitting their first draft and evaluate their own performance against the same criteria after completing the final product. Self-assessment helps students distinguish their own decisions from AI suggestions. Instead of asking students for a lengthy explanation for each dimension, they can be asked to identify one strength and one area for improvement. This brief reflection becomes a natural part of the process record.

Feedback should not simply become a list of errors in the final text. Hyland and Hyland (2006) note that feedback on second-language writing should support the student in rethinking the text and developing writing strategies. In an AI-supported task, the teacher can also provide feedback on how the student evaluates the tool's suggestions. Instead of saying, "This word is wrong," questions such as "Why did you choose this phrase based on the target audience?" or "How did the suggested change affect the tone of the text?" guide the student toward the rationale behind their decision.

Limited student involvement in preparing the evaluation criteria can also be beneficial. In a B1 class, students can discuss together what characteristics a good opinion piece should have; in a B2 class, they can determine what to watch out for to ensure that cultural explanations do not become generalizations. The teacher maintains the core criteria, and students discuss the meaning of these criteria using sample texts. In this way, the rubric ceases to be a technical document seen for the first time at the end of the assignment. Students know from the outset what questions to ask as they examine the AI output.

Peer feedback can also be used, but the responsibility for verifying the AI output should not be delegated to peers. Peers can comment on the text's clarity, target audience, and communicative purpose. Verifying the accuracy of a linguistic or cultural claim requires a dictionary, a corpus, an institutional resource, or the teacher's opinion. The combined use of human and AI feedback allows students to compare different suggestions, and the final decision ultimately rests with the student.

Student Anxiety and Accessible Assessment

Oral explanations and supervised initial performances, while enhancing assessment security, may cause anxiety for some students. In particular, spontaneous speech in a foreign language may make a student with strong written proficiency appear to perform below their actual level. Nallaya et al. (2024) note that issues such as student anxiety, equity, and standardization for second-language speakers must be addressed during the design phase of oral assessments. Therefore, the oral section should not be administered as a surprise interrogation. The duration, general question types, and scoring criteria must be explained in advance.

To ensure accessibility, accommodations such as repeating the question once, allowing a brief thinking period, using visual cues, or dividing the task into two short sections may be implemented. These supports should not teach the content of the response but should enable the student to demonstrate the targeted skill. While using a fully pre-recorded presentation may seem practical for a student experiencing speaking anxiety, it eliminates evidence of spontaneous interaction. A more balanced solution is for the teacher to ask one or two follow-up questions after a brief prepared introduction.

Digital accessibility is equally important. The AI tool used may not be compatible with screen readers, its audio features may not function with the same quality in every language, or it may require a paid subscription. UNESCO's human-centered approach requires that technological innovation not create new forms of exclusion among students (Miao & Holmes, 2023). Teachers should prepare equivalent alternatives that measure the same learning outcomes for students who cannot access the tool or who do not wish to share their personal data. Assessment equity relies not on providing all students with the same tool, but on ensuring comparable conditions in which they can demonstrate the targeted competency.

Discussion and Conclusion

Generative AI has amplified a pre-existing problem in assessment within French language instruction. The product a student submits is not the same as the student's ability to create that product independently. Translation tools, online dictionaries, and help from others have long made this distinction difficult. Large language models, however, have made the authorship of the work and the structure being assessed even more ambiguous, as they combine the functions of idea generation, text composition, editing, and rewriting in a single platform. Therefore, the solution is not to

continue with old assignments in the same format and seek more powerful detection software.

From a language assessment perspective, a more defensible approach is to diversify the evidence. An independent initial performance demonstrates what the student can do without external support. A documented revision process reveals how the student utilizes feedback. Critical validation shows whether the student views the AI output as the correct answer and whether they take responsibility for their own decision. Oral explanation and presentation, on the other hand, test the understanding of the prepared product and the use of linguistic resources in a new context. When these pieces of evidence are used together, the assessment shifts from the surface features of the final text to the student's communicative actions.

The proposed four stages are not a mandatory, fixed system for every lesson. In an A2 class, the process may be limited to explaining two changes and a brief oral question. At the B2 level, source verification, target audience, and cultural representation can be addressed in greater detail. In a conversation class, independent interaction and presentation take center stage, while in a translation class, justifying alternatives may be given greater weight. The validity of the assessment stems not from the number of stages but from its alignment with learning outcomes.

A contribution specific to French language instruction is to avoid discussing the use of artificial intelligence only in terms of grammatical error correction. Form of address, linguistic register, target audience, mediation, and cultural explanation are fundamental components of communicative competence in French. Students should be able to reject a grammatically correct sentence suggested by AI if it does not fit the context, verify cultural knowledge, and reformulate the same message for different audiences. The Common European Framework of Reference for Languages' understanding of social actors and mediation provides a strong foundation for such tasks (Council of Europe, 2020).

This approach has limitations. Process records may be incomplete, oral interviews may cause anxiety, and teacher workload may increase in large classes. Access to technology is not equal. The provided rubric and four-stage process are not an empirically validated, standardized assessment tool. Future studies should examine the applicability of the Common European Framework of Reference for Languages at different levels, inter-rater reliability, student and teacher perceptions, and the relationship between independent performance and AI-supported output. These limitations indicate that the proposal must be adapted to the specific context.

The quality of implementation depends more on task design than on the brand or version of the AI tool used. Tools can change rapidly; technical specifications valid today may differ tomorrow. In contrast, the principles of collecting independent performance, clarifying usage limits, verifying output against sources, and asking students to justify their decisions are more enduring. When teachers adopt these principles rather than following platform-specific prescriptions, it makes assessment design more resilient to technological change. Chuang and Yan's (2025) emphasis on sustainability and context sensitivity also underscores the need to avoid tool-centered solutions in language assessment.

In the age of generative AI, reliable French assessment cannot be reduced to simply asking students to prove they did not use the tool. A more meaningful question is what the student can do with French, how they manage the support they receive from the tool, and whether they can explain their final linguistic decisions and apply them in a new situation. Artificial intelligence is not the center of assessment; it is one of the conditions. The focus should be on the student's linguistic proficiency, communicative reasoning, and responsibility for learning.

*Artificial intelligence was utilised in this study to assist with language support and proofreading.

References

Chapelle, C. A. (2025). Generative AI as game changer: Implications for language education. System, 132, 103672. https://doi.org/10.1016/j.system.2025.103672

Chuang, P.L., & Yan, X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions. System, 134, 103846. https://doi.org/10.1016/j.system.2025.103846

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment—Companion volume. Council of Europe Publishing.

Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. Language Learning & Technology, 28(2), 5–31. https://doi.org/10.64152/10125/73570

Hamp-Lyons, L., & Condon, W. (2000). Assessing the portfolio: Principles for practice, theory, and research. Hampton Press.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. Language Teaching, 39(2), 83–101. https://doi.org/10.1017/S0261444806003399

Jang, E. E., & Sawaki, Y. (2025). Advancing language assessment for teaching and learning in the era of the artificial intelligence (AI) revolution: Promises and challenges. Language Testing, 42(4), 361–368. https://doi.org/10.1177/02655322251348685

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

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates. (2023). Assessment reform for the age of artificial intelligence. Tertiary Education Quality and Standards Agency.

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

Nallaya, S., Gentili, S., Weeks, S., & Baldock, K. (2024). The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review. Issues in Educational Research, 34(2), 629–646.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. Journal of University Teaching and Learning Practice, 21(6). https://doi.org/10.53761/q3azde36

Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. Studies in Higher Education, 45(11), 2132–2148. https://doi.org/10.1080/03075079.2019.1582015

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. Journal of Second Language Writing, 67, 101187. https://doi.org/10.1016/j.jslw.2025.101187

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, 26. https://doi.org/10.1007/s40979-023-00146-z

Yılmaz Güngör, Z. (2020). Fransızca Yabancı Dil Öğrencilerinin Dil Öğrenmeye Yönelik İnançlarının Çeşitli Değişkenler Bağlamında İncelenmesi. *Türk Eğitim Bilimleri Dergisi*, *18*(2), 571-594. https://doi.org/10.37217/tebd.691562

Living Portfolio in Iterative Artificial Intelligence Feedback and the Reorientation of Foreign Language Assessment

Alireza Navidmoghaddam [1]

Abstract
Artificial intelligence is entering foreign language classrooms faster than research can synthesise it, leaving an evidence base that is empirically active but conceptually fragmented: writing gains are studied apart from motivational gains, biometric stress markers apart from grammatical accuracy, portfolio design apart from corpus-based developmental tracking. This chapter offers an integrative conceptual reading of that literature, drawing on seven empirical studies read in full—spanning AI-enhanced portfolio writing, AI-assisted portfolio assessment, biometric-informed adaptive reading, a longitudinal speaking corpus, task-based instruction, and a review of learner emotion—alongside reviews and framework proposals. It argues one mechanism recurs across unrelated designs: sustained, proficiency-calibrated feedback embedded in an iterative, documented cycle moves learners along linguistic, cognitive, and affective lines simultaneously, unlike isolated single-session interventions. This is treated as a theoretically motivated proposition, not an established finding, with explicit falsification conditions. The chapter also specifies boundaries: pragmatic competence resists automated feedback for reasons that appear structural rather than transitional; data governance obligations grow as portfolios accumulate longitudinal records; and the evidence base is narrower—geographically, institutionally, bibliometrically—than surrounding rhetoric implies. It closes by outlining implications for portfolios as living instruments and a teacherly role centered on judgement and relationship, proposing a research agenda to test that direction.
Keywords: AI-enhanced portfolios; second language development; emotion regulation; longitudinal assessment; feedback literacy.

Introduction

In no time, artificial intelligence (AI) has become one of the driving forces that redefine how foreign languages are taught and learned. While chatbots are capable of correcting grammar errors at the moment of speaking, adaptive platforms can modify the difficulty of the reading material according to learners' fixation behaviors, and portfolio systems can provide personalized feedback on drafts being written. Recent systematic reviews and bibliometric studies concur about the rapid but uneven evolution of this empirical literature, as it has focused too much on second language

[1] Lecturer Dr., Istanbul University-Cerrahpaşa, School of Foreign Languages, Department of Foreign Languages, English Preparatory Program, ORCID: 0000-0002-8804-6761, *alirezan@iuc.edu.tr*, Türkiye

(L2) writing and not nearly enough on speaking, listening, and many more relevant pedagogical and ethical issues raised by the technology (Li et al., 2025; Wang et al., 2026). For instance, bibliometric mapping of the 2017-2023 period shows the same rapid increase in the amount of publications along with the relentless concentration of works within few skills areas and countries (Rahman et al., 2024), while a review of reviews that analyzes the years 2024-2025 suggests that even synthesis literature begins to reproduce itself faster than its consolidation (Jeon, 2025).A parallel systematic review of 37 empirical studies on learner emotion in AI-assisted English as a second or foreign language (ESL/EFL) learning confirms a similar imbalance: anxiety and enjoyment dominate the literature, while boredom, interest, shyness, and the mechanisms of emotion regulation itself remain comparatively neglected (Zhang & Liu, 2025).

The result is a body of knowledge that is individually rigorous but collectively fragmented. A study of AI-enhanced e-portfolios reports on writing proficiency and self-efficacy (Xiao & Liu, 2025). A study of biometric-informed adaptive reading reports on comprehension, anxiety, and cognitive load (Yuan, 2025). A study of AI-assisted portfolio assessment reports on mindfulness and academic emotion regulation (Khasawneh et al., 2025). A longitudinal, error-annotated speaking corpus reports on grammatical accuracy trajectories and self-correction (Liu et al., 2026). Reviews focused specifically on L2 writing (Luo et al., 2025) and on generative AI for written corrective feedback (Ding et al., 2025) catalogue these findings skill by skill. Each result is real and specific. What is largely absent is an account of how they relate to one another and of what, taken together, they imply about where foreign language pedagogy may be heading rather than merely where it currently stands.

This chapter attempts that integrative and prospective account. Rather than cataloguing the literature intervention by intervention, it reads the evidence thematically, allowing findings on writing, reading, speaking, and emotion to inform one another in the way a practising teacher or programme designer would need to use them. The argument proceeds in four moves. The Theoretical Framework section sets out the theoretical framework through which the evidence is read. The sections that follow—Instructional Infrastructure or Discrete Tool, From Isolated Correction to Iterative Documentation, Cognition and Emotion, and Development Over Time—identify a recurrent mechanism—iterative, proficiency-calibrated, learner-visible feedback embedded in a documented cycle—across four otherwise unrelated bodies of evidence. The Boundaries of the Current Evidence section specifies the boundaries of what that evidence can support. The Implications section draws out implications for practice and

proposes a testable research agenda. The central claim is deliberately structural rather than technological: what appears to do the pedagogical work is the architecture of the feedback loop, not any particular platform, vendor, or model generation.

Scope, Source Selection, and Method of Synthesis

Because the credibility of an integrative argument depends on the transparency of its evidence base, the scope of this chapter is stated explicitly before the argument begins. According to Torraco (2005) and Whittemore and Knafl (2005), this is an integrative conceptual review on its own. It seeks to produce a theoretical explanation from diverse sources of empirical and conceptual work rather than compute an aggregated effect. Consequently, it is not a systematic review, nor does it claim completeness or replicability of search associated with the systematic review. Thus, while evaluating it in terms of narrative-review criteria (Baethge et al., 2019), readers should take into consideration the conclusions made.

For the purpose of close reading, five primary studies were chosen according to three criteria; they must be recent (published in 2025 or later); they must feature either iterative feedback or longitudinal feedback instead of a one-time intervention; and in total they should cover all skills, productive and receptive ones, as well as the affective dimension. During writing, two additional studies were included since they offered an independent test of one of the emerging claims: task-based language teaching quasi-experiment that expanded the approach to speaking and reading (Huang et al., 2024) as well as a correlational study of AI use in all four skills in a portfolio approach (Le et al., 2026). All seven were read in full rather than at abstract level, so that every claim attributed to them here can be checked against the original reporting of methods, statistics, and findings. Table 1 summarises their designs and principal outcomes.

A wider set of systematic reviews, bibliometric analyses, and framework proposals is cited more selectively, to characterise the pace and shape of the field's growth and to situate the primary studies within it, rather than to supply additional empirical claims. Foundational theoretical sources are cited where the argument depends on a construct they define.

Three limitations of this selection procedure should be recorded at the outset. First, the criterion privileging iterative and longitudinal designs is not neutral with respect to the chapter's conclusion: a set of studies chosen partly because they instantiate sustained feedback cycles is more likely to support a claim about the value of sustained feedback cycles. The

convergence reported below is therefore best understood as an internally consistent reading of a purposively assembled corpus, not as an unbiased estimate of the field. The What Convergence Can and Cannot Establish subsection returns to what such convergence can and cannot establish. Second, no formal risk-of-bias appraisal was conducted, and the sources vary considerably in methodological transparency and in the editorial rigour of their outlets; where a claim rests on a single low-visibility source, this is noted in the text. Third, the corpus is small by design, which permits close reading at the cost of coverage. The Research Agenda subsection specifies the studies that would be needed to test the resulting propositions against a broader base.

Table 1: *Designs and principal outcomes of the studies read in full*

Study	**Design, setting, and duration**	**Primary focus**	**Principal reported outcome (see original sources for exact statistics)**
Xiao & Liu (2025)	Quasi-experiment; junior-high EFL writing instruction; 12 weeks; AI-enhanced e-portfolio vs. pen-and-paper portfolio (N = 40: 21 experimental, 19 control)	L2 writing; writing self-efficacy	Larger gains in writing proficiency and self-efficacy in the AI-supported condition. Writing scores: experimental M = 40.82 (SD = 12.42) pre-test to M = 54.45 (SD = 13.31) post-test; control M = 41.20 (SD = 14.31) to M = 46.72 (SD = 15.20); ANCOVA $F(1, 38) = 18.12$, $p = .003$, partial $\eta^2 = .236$. Self-efficacy (post-test total): experimental M = 4.82 (SD = 0.29) vs. control M = 3.64 (SD = 0.52), $t = 9.90$, $p < .001$.
Khasawneh et al. (2025)	Intervention study; higher-education EFL; AI-assisted portfolio assessment	Emotion regulation; mindfulness; learning attitudes	Gains in academic emotion regulation, mindfulness, and attitudes attributed to continuous individualised feedback (N = 69: 38 experimental, 31 control; quasi-experimental, ANCOVA/t-tests).
Yuan (2025)	Mixed-methods; adaptive AI reading platform with heart-rate variability	Reading comprehension; anxiety;	Comprehension gains with concurrent reductions in reported anxiety and cognitive load; qualitative accounts

Study	Design, setting, and duration	Primary focus	Principal reported outcome (see original sources for exact statistics)
	sensing and eye tracking vs. static digital texts	cognitive load	include unease about biometric data (N = 300, randomly assigned; effect sizes ranged from moderate to large).
Liu et al. (2026)	Longitudinal, error-annotated learner corpus; Chinese EFL spoken English; two semesters, three measurement points; AI-enhanced smart classroom (N = 100: 28 lower-intermediate, 47 upper-intermediate, 25 advanced)	L2 speaking accuracy; self-correction; developmental trajectories	Self-correction rate rose from 12% (T1) to 34% (T3), correlating with proficiency gains ($r = .52$, $p < .001$); cluster analysis identified three developmental profiles: Rapid Developers (31% of participants), Steady Progressors (52%), and Plateau Learners (17%); pragmatic accuracy improved only marginally (23-31% error reduction) compared to morphosyntactic domains (68-76% reduction).
Zhang & Liu (2025)	Systematic review of 37 empirical studies	Learner emotion in AI-assisted ESL/EFL learning	Anxiety and enjoyment dominate the literature; boredom, interest, shyness, and emotion-regulation mechanisms under-researched.
Huang et al. (2024)	Quasi-experiment; college English; 12 weeks; AI-supported task-based language teaching vs. conventional TBLT (N = 120: 60 experimental, 60 control)	Speaking; reading comprehension; engagement	Gains in speaking and reading comprehension and higher engagement; instructors report infrastructure and training costs. Speaking scores rose significantly in the experimental group ($t = 4.25$, $p < .01$); reading comprehension also improved significantly ($t = 3.87$, $p < .01$); engagement was significantly higher in the experimental group (ANOVA $F(1, 118) = 5.67$, $p < .05$).

Study	Design, setting, and duration	Primary focus	Principal reported outcome (see original sources for exact statistics)
Le et al. (2026)	Correlational; Vietnamese non-English majors; portfolio-based General English; official end-of-course examination scores (N = 107)	Frequency of AI use across the four skills	AI use frequency correlated significantly with English proficiency: GPA r = .630, reading r = .962, writing r = .821, speaking r = .557 (all p < .01), and listening r = .249 (p < .01, comparatively weak). The near-perfect reading coefficient requires cautious interpretation, given close alignment between AI-supported reading tasks and the exam format.

Note: TBLT refers to task-based language teaching. Statistics reported in the table are drawn directly from each primary source; readers should consult the original sources for full statistical detail (e.g., complete ANCOVA and post-hoc tables).

This study did not require ethics committee approval, as it is a conceptual review that synthesises published literature rather than collecting original data from human or animal participants.

Theoretical Framework

The mechanism this chapter identifies is not a novel theoretical construct. It is better understood as a technologically enabled instance of four established positions, and stating them explicitly makes the argument falsifiable rather than merely plausible. First, feedback is effective to the extent that it addresses the gap between current and desired performance at a level the learner can act upon, and ineffective when it is delivered too late, too generically, or at the wrong level of abstraction (Hattie & Timperley, 2007). This predicts that automated feedback should help most where it is immediate, specific, and calibrated—and least where the relevant gap is not one of form.

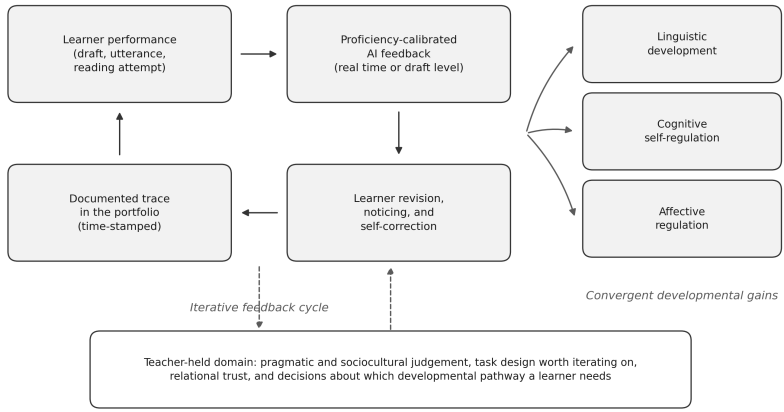
Second, learning gains accrue when instruction operates within the zone of proximal development, the band in which a task exceeds independent capacity but is achievable with support (Vygotsky, 1978). Adaptive calibration of difficulty is a mechanised approximation of that principle, and it predicts that support arriving during a task will outperform support

arriving after it—a prediction the Pragmatic Competence subsection tests against the listening data.

Third, iterative revision is a self-regulatory activity rather than a corrective one: what develops across cycles is the learner's capacity to monitor and evaluate their own production (Zimmerman, 2002), and the uptake of feedback depends on a learner's developing feedback literacy rather than on the quality of the feedback alone (Carless & Boud, 2018). Evidence from tertiary writing instruction indicates that iterative writing with feedback changes the characteristics of learners' texts, not merely their scores (Vardi, 2011). If the mechanism proposed here is correct, the durable outcome of AI feedback should be improved self-monitoring, which is measurable independently of the AI system's presence.

Fourth, affective and cognitive development are not separable channels. Control-value theory holds that achievement emotions arise from learners' appraisals of control over, and value of, an activity (Pekrun, 2006); foreign language anxiety is a situation-specific construct responsive to precisely such appraisals (Horwitz et al., 1986); and cognitive load theory predicts that reducing extraneous processing demands frees capacity for schema construction (Sweller, 1988). Feedback that makes progress visible and keeps difficulty tractable should therefore raise perceived control and lower anxiety through the same operation that produces linguistic gain. This is the theoretical reason to expect the affective and linguistic findings reviewed below to travel together, and it is what distinguishes the present argument from the weaker claim that AI happens to be associated with several desirable outcomes at once. Figure 1 represents the resulting mechanism schematically, including the domain the evidence assigns to human instruction.

Figure 1: *The iterative feedback cycle and its proposed developmental outcomes*



Note. The four-phase cycle on the left represents the mechanism identified across the studies reviewed from From Isolated Correction to Iterative Documentation through Development Over Time. The dashed arrows indicate that teacher judgement both shapes the tasks entering the cycle and interprets the documented trace it produces. The figure represents a proposed mechanism rather than an empirically isolated causal pathway; see the What Convergence Can and Cannot Establish subsection.

Instructional Infrastructure or Discrete Tool

It is worthwhile to reconsider what we mean by the term "AI in language education." Originally, AI was understood as a certain type of tool—such as a grammar checker or a chatbot—and the debates focused on whether such AI tools were efficient or not (Özçelik & Ekşi, 2024). Nowadays, we talk about AI in terms of infrastructures—the base of technology behind a certain course, which can be configured in various ways depending on the objectives of the course. It is important to note that things like automatic speech recognition, feedback from a large-language model, biometric tracking, and portfolio technologies should not be treated as separate competing technologies but as components combined together—and the current research shows that the best results are achieved when these components are combined rather than applied separately (Nouzri et al., 2024).

The transition discussed is of importance since it alters the way the field should be interpreted. A chatbot serves simply as a conversational tool

while different analyses exploring the potential of chatbot-assisted language learning demonstrate that different benefits brought by the tool are very limited and quite temporary (Huang et al., 2021). A portfolio serves solely for documenting purposes, with the overall evidence indicating that documentation, without adequate feedback, serves only as a formality rather than an effective tool (Abdulhamid & Fox, 2020). However, the point is that when both tools are combined and appropriately utilized according to the learner's proficiency, they perform their function. This distinction—feedback as recurring process versus feedback as isolated correction—runs through nearly every strand of evidence considered below, and is the thread the remainder of the chapter follows.

The structural assessment justifies the rapid and uneven progress of the literature. When a technology is seen as merely a tool, the research question becomes simple: works it? When it is seen as infrastructure allowing several configurations, the questions burgeon: which configuration, for what skill, what proficiency level, which population, for how long—and the answers come more often than any synthesis has time to unpack them (Jeon, 2025). The opposite of synthesis does not yield a more coherent picture of the evidence but more pieces of evidence besides those already.

From Isolated Correction to Iterative Documentation

Evidence from AI-enhanced e-portfolios

The clearest real-world evidence for this change has come from a comparison of AI-driven e-portfolios with the standard pen-and-paper portfolios in junior-high EFL writing classes over the period of twelve weeks (Xiao & Liu, 2025). The AI-based group showed a greater increase in both the level of proficiency and self-efficacy than their counterparts. Nonetheless, the most interesting thing is not whether the innovation worked but how exactly it was done. Students submitted drafts, were given grammar, coherence, and style feedback, then revised their texts and, only after that, received the teacher's feedback again. All these stages are essential for the ability of students to see the changes in their own papers according to specific and calculated input. The point that section 3 is making is that what the innovation did is not simply correcting the errors—both options involve correction—but a combination of the ability to see the process of revision from start to finish. The fact that self-efficacy increases with increasing proficiency supports the control appraisal theory of Bandura (1997) and Pekrun (2006), as opposed to any purely linguistic explanation of the phenomenon.

Collapsing the distance between drafting and feedback

Technical literature on the topic raises questions regarding the ways in which these cycles can be integrated into already established learner systems instead of asking users to leave an e-portfolio for a separate feedback application. For instance, Osawa (2023) discusses the use of e-portfolios and proposes delivering automated written corrective feedback in the drafting environment of the e-portfolio by utilizing a universal productivity app and its AI functions based on the logic that feedback-rich portfolios are already self-regulating, and that bringing drafting closer to feedback will make it more effective as such. This idea is a theory, not one that has been proven. This indicates that it should be treated as such. The exception is that the future will bring some changes in its application: as portfolio systems and AI feedback systems mix up and become one platform, the current barriers to revision will cease to exist, making the process of revision much more efficient. However, the question about the effects of fewer barriers on revision frequency remains unanswered (see the Research Agenda subsection).

Convergence with an independently developed instructional framework

A pedagogical framework for AI-mediated L2 writing instruction developed around active, inquiry-based, and adaptive learning rather than around portfolios reaches a compatible conclusion from a different starting point: learners instructed under the framework improved in both writing motivation and writing outcomes (Wang et al., 2025). A specific case illustrates the application of this concept outside the realm of writing and the setting of portfolio construction. A quasi-experimental study conducted for a duration of 12 weeks explored the use of AI tools in TBLT in a university English course and discovered that it was more effective in terms of improvements in speaking and reading skills than traditional TBLT, and had higher levels of engagement while running (Huang et al., 2024). The reasoning behind teaching was similar to the logic applied earlier—AI-based applications provided students with personalized pre-class preparation and real-time feedback during their in-class activities turning a prescribed sequence of tasks into a repeated and personalized sequence of activities. The qualitative results introduced a word of caution elaborated on in the Geographic, Institutional, and Bibliometric Concentration subsection, as both teachers and students have noticed that successful implementation depends on having the necessary infrastructure, as well as on the amount of time spent learning to use the tools, and both groups

stressed the fact that support should be provided by the institution instead of arguing that the technology can function without help of the institution.

From the perspective of feedback design rather than empirical evaluation, a conceptual counterpart arrives at a structurally similar prescription. Zou (2025) suggests the development of a hierarchized feedback landscape including AI feedback addressing low-order concerns such as grammar, syntax, and vocabulary; peer feedback aimed at coherence and cohesion; and teacher feedback concerned with conceptualization, argumentation, and discipline-specific judgment. Each layer compensates for a certain drawback of its predecessor--while AI feedback has immeasurable and constant nature, it lacks context knowledge, peer feedback is collaborative yet of variable quality, and teacher feedback, although focused, is limited by the scarcity of time and efforts. The same discussion extends the case for portfolio-based, multi-draft assessment from a further angle, arguing that reflective journals, revision histories, and multimodal artefacts are comparatively resistant to the assessment-integrity concerns accompanying generative AI in higher education, precisely because they document developing thinking over time rather than a single finished output that could have been produced in one pass.

What convergence can and cannot establish

The observation at the heart of this chapter is that a shared design principle appears again and again in these sources, leading to a particular objection. The apprehension is that in trying to ascertain commonality across disparate studies one runs the risk of treating different forms of evidence as if they were the same. In order to address this issue, three different caveats need to be put forward.

First, independence is partial. Xiao and Liu (2025), Wang et al. (2025), and Huang et al. (2024) are separate empirical studies with different populations, instruments, and theoretical starting points, and their agreement is evidentially meaningful. Osawa (2023) and Zou (2025), by contrast, are conceptual proposals; their agreement with the empirical studies reflects a shared reading of the same prior literature at least as much as it reflects independent corroboration. Conceptual convergence and empirical replication should not be counted as the same kind of evidence, and this chapter does not claim otherwise.

Second, the corpus was purposively selected for iterative designs (see the Scope, Source Selection, and Method of Synthesis section, above), so the convergence is in part an artefact of selection. What the studies collectively establish is that the mechanism is *consistent with* results obtained

under varied conditions, not that it has been isolated. No study reviewed here manipulates iteration while holding feedback quality constant, which is the design that would identify the mechanism directly.

Third, and most importantly, the claim is falsifiable and should be stated so. If iteration and learner visibility are doing the pedagogical work, then a high-quality single-shot feedback condition should underperform a lower-quality iterative condition matched for total feedback volume; learners' self-monitoring should improve more than their reliance on the tool; and gains should be largest where the target competence is form-based and observable. The Pragmatic Competence subsection shows that the third prediction holds in at least one instance. The first two remain untested, and until they are tested the mechanism proposed here is best treated as a theoretically motivated hypothesis with convergent circumstantial support—a stronger position than an isolated finding, but weaker than a demonstrated causal claim.

Cognition and Emotion

If writing research indicates what iterative AI feedback does for linguistic production, a second strand indicates what the same logic does when extended to learners' physiological and affective states. An investigation employing mixed methods on a reading platform that integrates AI technology with heart-rate variability and eye tracking, as well as normal comprehension tests, mentions that the reading platform produced better comprehension with less anxiety and cognitive load than static resources with the same content (Yuan, 2025). However, the process relied on a method used in writing literature, with the difference that this time it makes adjustments in the real time. Thus, when the analysis of the reading patterns shows any difficulties for the learners, the reading platform simplifies the text or provides a brief explanation instantly, keeping students in the zone of possible development (Vygotsky, 1978) and preventing frustration.

The qualitative accounts add a dimension the aggregate statistics do not. Learners described the platform as engaging and as reducing their reading anxiety, and many valued the adaptive features and the interface. Learners also attended to their own biometric readings during the process, with reported improvements in reading strategies and self-awareness. Not every response was positive: a minority found the biometric display distracting, and some expressed discomfort about how their physiological data were collected and stored—a concern the Data Governance and Learner Consent subsection develops.

Two construct-validity caveats should temper the interpretation of this study, and by extension of any biometrically instrumented design. Heart-rate variability is a peripheral index of autonomic arousal, not a measure of anxiety as the construct is defined in the foreign language anxiety literature (Horwitz et al., 1986); arousal attributable to interest and arousal attributable to threat are not readily distinguishable from the signal alone. Similarly, regressive fixations are an indirect proxy for processing difficulty, and can also reflect careful reading. Where a study reports concurrent movement in a self-report anxiety scale, a physiological index, and a comprehension score, the convergence is informative, but the inference that the physiological signal measured the affective construct remains an interpretation rather than a demonstration. Reporting the association between the biometric and self-report measures directly would strengthen such claims considerably.

A related line of evidence indicates that AI-assisted portfolio assessment is associated with gains in mindfulness and academic emotion regulation, again through continuous individualised feedback rather than one-off assessment events (Khasawneh et al., 2025). Analysing learners' written reflections for indications of anxiety, stress, or positive engagement allowed the system to surface affective information that a human reviewer working through a stack of submissions might miss, and to do so consistently across a cohort rather than only for learners who sought additional support. Read alongside the systematic-review evidence on learner emotion in AI-assisted ESL/EFL learning—where enjoyment mediates the relationship between self-efficacy and continued tool use, and where AI-engaged learners report decreased anxiety alongside increased willingness to communicate (Zhang & Liu, 2025)—a consistent picture emerges. AI's affective correlates in language learning do not appear to be a by-product of its cognitive correlates; both appear to operate through the same continuous-feedback structure, as control-value theory would predict (Pekrun, 2006). The causal direction, however, is not established by any of these designs: reduced anxiety may enable comprehension, comprehension gains may reduce anxiety, and a third factor such as engagement may drive both.

Development Over Time

The previous paragraphs illustrate development over weeks. An error-tagged longitudinal corpus of Chinese students' spoken English in an EFL context, generated through three measurements across two semesters with the use of technological advancements in classroom teaching, adds great insights into the case (Liu et al., 2026).

There are two results that should be noted. First, the degree of spontaneous self-correction of learners increased during the duration of the experiment, which correlated with the total improvement in language proficiency—suggesting the indication that AI feedback is beneficial in developing self-monitoring, the self-regulatory outcome suggested in the Theoretical Framework section (Zimmerman, 2002). The design cannot establish that the AI feedback caused the increase, since self-correction also rises with proficiency generally; a condition in which the AI is withdrawn would be required to show that the capacity persists in its absence. Second, cluster analysis of individual trajectories identified three developmental profiles—rapid developers, steady progressors, and learners who plateaued after early gains—that were not visible in the aggregate statistics and that were associated with digital literacy, intrinsic motivation, and voluntary engagement with the tools outside required class time. A single-point assessment would have missed this heterogeneity entirely. Cluster solutions are, however, sensitive to algorithm and to the number of clusters specified, and a three-profile solution should be treated as a description of this cohort pending replication.

This is, in effect, the argument for portfolio-based longitudinal tracking made concrete through corpus methodology rather than through portfolio architecture. Where the writing and reading literatures show that iterative feedback is associated with gains within a single intervention (Vardi, 2011), the speaking corpus shows what becomes visible only when iteration is tracked across an extended period: not a single average effect but a landscape of developmental pathways that a responsive pedagogy would need to recognise individually.

Unlike other studies that are based exclusively on self-report measures, the correlational examination of AI in portfolio-based instruction in General English among non-major students in Vietnam allows to examine one aspect of this correlation. This research shows, based on results from standardized tests, that there is a strong link between usage of AI and effectiveness of reading and writing, moderate connection with speaking, and practically no connection with listening (Le et al., 2026). Thus, we can once again observe the findings of error-annotated corpora concerning the discrepancies between the development of accuracy and pragmatic competency, although it is done with a different method and group of participants (Liu et al., 2026). AI feedback has a significant impact on text-related tasks where it deals with grammar correction, vocabulary suggestion, and paragraph restructuring. However, it loses its effectiveness when it comes to the processing of real-time, socially motivated, or purely auditory information. Two independent studies conducted on two continents

have arrived at the same type of asymmetry, thereby reinforcing the general hypothesis about the strengths and limits of AI in this respect but with the exception noted in the Statistical and Measurement Caution subsection.

Boundaries of the Current Evidence

Pragmatic competence

Pragmatic appropriateness, which refers to the understanding of what is socially appropriate in a certain context as opposed to what is grammatically accurate, manifested only slight improvement in the longitudinal speaking corpus while morphosyntactic accuracy had been significantly improved, with the learners continuing to generate overly direct and situationally inappropriate language long after they had resolved other types of errors (Liu et al. 2026). Based on the framework in the Theoretical Framework section above, there is nothing surprising about this. In order to be effective, feedback has to show the actionable gap (Hattie & Timperley 2007). Pragmatic competence is related to being cognizant of socially situated processes regarding interlocutors, context, etc. This gap can not be revealed in the surface structure of the utterance. Whether it is a question of system limitations or it just characteristic of this exact type of systems is an empirical question that should be asked rather than taken for granted. As far as education institutions and programme developers are concerned, it is better to start planning human-oriented instruction in pragmatics instead of relying on IT-based solutions to solve this issue.

Statistical and measurement caution

The study of Vietnamese portfolio indicates a strong relationship ranging from AI utilization and reading results having a correlation of $r = .962$ (Le et al., 2026)—this coefficient means that about 93% of the variance in reading results is accounted for by a single self-reported behavioral frequency. Based on this statistic, studies in educational research of complex outcomes with such high levels of association usually refer to some measurement error rather than to the substantive impact. Plausible sources include common-method variance where predictor and criterion are collected in proximity; restriction or coarse categorisation of the predictor, which can inflate coefficients computed on aggregated or ordinal data; construct overlap, if frequent AI users are also the students whose coursework most closely resembles the examination format; and small-sample instability. None of these can be adjudicated from the published report alone. The methodological point generalises: correlations at this

magnitude in this literature warrant scrutiny of what was measured before they are cited as evidence of what AI achieves, and the present chapter treats the study's *ordinal* finding—reading and writing above speaking above listening—as its informative contribution, not the absolute magnitudes.

The same study's weakest result is arguably its most useful. Listening, the one skill where AI use showed no meaningful association with improvement, was also the skill for which students used AI post hoc, generating transcripts to check after they had already worked through an audio passage, rather than as scaffolding embedded in the listening process itself. Read against the zone-of-proximal-development account (Vygotsky, 1978), this is the predicted outcome: scaffolding that arrives after a task cannot support performance during it, and the adaptive-reading evidence indicates that real-time calibration rather than after-the-fact correction is what accompanies comprehension gains (Yuan, 2025). The apparent failure of AI to support listening development may therefore reflect a limitation of deployment rather than of the technology—a distinction that is testable, and that the Research Agenda subsection proposes as a research priority.

Data governance and learner consent

Systems generating longitudinal, granular records of learner writing, speech, physiological response, and affective self-report accumulate a substantial quantity of sensitive information, and the evidence indicates that learner comfort cannot be assumed even where formal consent procedures are followed correctly: a minority of participants in the adaptive-reading study expressed unease about the collection and storage of their physiological data (Yuan, 2025), and research on learner–instructor interaction in AI-mediated online learning documents comparable concerns about surveillance and the erosion of interpersonal trust (Seo et al., 2021). As portfolio systems accumulate data across multiple semesters rather than single studies, and as biometric and behavioural logging becomes more common, these concerns are likely to grow rather than recede. Three criteria emerge in a straightforward manner: the data collection should be reduced to the level necessary for achieving the teaching goal, the duration of keeping the data and access to it should be determined before the data collection and not afterward, and participation in the biometrics should be truly voluntary because when people give consent under evaluation conditions it means they are giving consent due to pressure put on them by an organization. Transparency in data governance is perceived correctly as an important design criterion and not as some administrative aspect.

Geographic, institutional, and bibliometric concentration

This report puts together its main evidence base and mainly draws on East and Southeast Asian EFL backgrounds, mainly Chinese secondary and tertiary contexts with the addition of one higher education study carried out in Vietnam and another study conducted somewhere else, which means that the intervention studies adopted a single-institution research design. In fact, this type of sampling makes it clear that it limits the generalisability of applied linguistics studies even more than it is normally pointed out (Andringa & Godfroid, 2020). Not surprisingly, the conclusions are true from the perspective of the findings but the idea of "AI-enhanced language learning" is based on a much narrower cultural and institutional background than one may expect when hearing about this concept.

Another seldom mentioned issue is the bibliometric one. The sources summarized in this section have been selected from the recognized journals through to conference proceedings and obscure regional journals which make it hard to establish the standards of the peer-review process just based on information about it. Since the field is expanding faster than its review capacity (Jeon, 2025; Rahman et al., 2024), the proportion of rapidly published, lightly reviewed work in any current synthesis is high. The claims made in this chapter should be assessed partly depending on the venue of study on which they are based, and the purpose of the Table 1 summary is not to show how each study compares in terms of evidential weight.

An equity issue that has emerged based on theory rather than practice is the possibility of having further reliance on technology in education leading to the increase of educational inequality: as AI tools for feedback and personalization grow more advanced, there is a chance that only wealthy institutions would benefit from them and use them on a large scale; this point also raises a concern regarding how convenient access to technology might lead to an overdependence, with learners failing to spend enough time on planning, revising and presenting work in language education (Zou, 2025). Parallel to this concern, teacher education research has concluded that teacher training should be improved to help educators adapt to this new reality rather than relying on them acquiring skills in an informal way (Karakaya et al., 2025). In line with the current approach, the TBLT study demonstrates that there are practical issues as well: instructors expressed their frustration with the need to allocate time to learn how to use new instruments and requested assistance from their institutions (Huang et al., 2024).

Technological churn and research tempo

A final, structural limitation is the pace at which the underlying technology changes relative to the pace at which longitudinal research can be conducted and published. Every primary study synthesised here was designed around a specific generation of AI tools, over windows ranging from a single semester to two academic years; by publication, the tools evaluated have often been superseded by systems with different capabilities. This is not a flaw in any individual study but a structural feature of researching a fast-moving technology with methods—quasi-experiments, longitudinal corpora, systematic reviews—developed for a slower-moving one. It counsels a particular form of humility for the prospective argument that follows: the specific findings reviewed here are likely to be superseded in their particulars well before the mechanism they point towards is. The safest forward-looking claims are therefore claims about mechanism and structure rather than about any specific tool, platform, or reported effect size, and the Implications section is written with that distinction deliberately in mind.

Implications

Taken together, and subject to the boundaries set out above, this evidence points towards a specific direction for foreign language pedagogy—not simply more AI, but a change in what a language-learning record is for. For most of the history of language teaching, assessment has been organised around discrete moments: the test, the essay, the oral examination, each producing a single score standing in for a much longer developmental process (Uyanıker, 2017). The evidence reviewed here suggests something different is now practical: assessment and instruction that treat development itself, over time, as the primary object of attention, with portfolios functioning not as archives of finished work but as living instruments tracking a trajectory as it forms.

Three projected shifts

Three shifts follow from the evidence. They are extrapolations from a narrow base and should be read as a plausible direction of travel rather than an established trajectory.

First, feedback is likely to move away from single-point correction towards continuous, proficiency-calibrated cycles embedded in the tools learners already use for drafting and practice—the trajectory visible in the movement from standalone automated written corrective feedback tools towards feedback embedded natively within portfolio interfaces (Osawa,

2023), and in frameworks that build iteration into instructional design rather than treating it as an optional extra (Wang et al., 2025).

Moreover, it is well-expected that the focus of assessment will cease to be strictly confined to linguistic precision, and start involving cognitive and affective aspects which have formally been neglected for a long time—awareness, anxiety level, motivation, self-control—not because the latter have become important but, because the tools powered by AI facilitate the measurement of these aspects on a large scale and for extended periods of time (Khasawneh et al., 2025; Yuan, 2025). The warnings provided in the Cognition and Emotion section and the Data Governance and Learner Consent subsection should be also taken into account: the fact that something is measured does not guarantee its validity, and the extent of new measurements is proportional to the responsibility imposed on it.

In addition, tracking over an extended period is expected to demonstrate as the evidence from the speaking corpus does that there does not exist a single typical learner trajectory that the instruction should be built around. Rather, there are a limited number of different developmental trajectories for which different instructional support should be provided rather than a typical curriculum used in a standardized manner (Liu et al., 2026).

The redistribution of teacherly work

None of this displaces the teacher; if anything, it clarifies what teachers are for. With the continued advancement of AI systems when it comes to providing real-time feedback that is tailored and focused, the remaining utility of human teachers is found in those areas of human-specialised knowledge where AI has clear limitations, namely practical and sociocultural judgement, the creation of activities that require the feedback process to begin, and human contact and trust that biometric personalisation will never replace. In a way, to come to this conclusion means to confirm the validity of the model of education as put forward by Zou (2025), whose model assigns the role of educators in assigning meaning, shaping arguments, and determining the scope of knowledge to be supplied by students in the process of education, while lower order inquiries would be completed both by the AI and the peers of respective students. Despite the fact that such redistribution is not inherent, it does call for the safeguarding of the released time so that it can be spent on higher functions instead of the administrative burden of the systems, and it calls for teacher training in preparation for the new role of the instructors (Karakaya et al., 2025).

Practice, and Presupposes

To elaborate further, a learner portfolio does not get opened only occasionally to obtain a score, while it collects evidence on a continual basis such as drafts and edits, audio recordings and corrections made by the learner themselves, entries reflecting their thoughts along with emotional signals transmitted within the entries and the entire portfolio setup in such a way that it represents a particular rich information flow instead of a final product represented by a single score. Thus, the teacher will be able to find not a score per se in the learner portfolio but a picture of the learner's development process and achievement levels established by the longitudinal study instead of understanding how the learner's abilities have been progressing (Liu et al., 2026). Curriculum design assumes of a shift from a standardized curriculum to creation of a number of the differentiated pathways adjusted to the learners' progress where the AI technology will control everything connected to adjustment of difficulty levels and providing feedback, while teachers will be responsible for decision making regarding selection of a particular pathway and its changes.

The requirements are significant and should be mentioned here. The vision calls for institutions ready to consider transparency in data governance not merely as an add-on to the process, but during its design phase, in view of learner anxieties with respect to biometrics and behavioral data collection even when provided proper consent (Seo et al., 2021; Yuan, 2025). It requires teacher training programs ready to see AI literacy as a fundamental competency rather than a supplementary workshop (Karakaya et al., 2025). It necessitates accounting for equity challenges in the sense that there is a risk for capable systems to become useful only for well-funded institutions (Zou, 2025). Besides, it calls for scholars willing to experiment with the boundaries discussed here rather than assume that they will disappear by themselves, specifically the boundary concerning pragmatic competence that can be the area that separates what a machine does teach from what can be taught only by a living person.

Research agenda

The claims presented here can be examined, and it is possible to devise studies to do so. Five major implications can be derived from the gaps discussed before. (a) Determining the impact of iteration: A study that controls for the overall volume and quality of feedback and tests different frequencies of giving and receiving feedback will allow checking whether iteration has a significant impact apart from feedback volume. (b) Transfer and withdrawal*:* Measuring self-monitoring and accuracy after AI support is removed would establish whether the self-correction gains reported

by Liu et al. (2026) reflect internalised capacity or tool dependence. (c) *Timing in receptive skills.* A direct comparison of in-task and post-task AI scaffolding for listening would test the deployment-versus-technology interpretation offered in the Statistical and Measurement Caution subsection. (d) *Construct validation of affective instrumentation.* Studies reporting biometric indices should report their association with validated self-report measures of the target construct, so that arousal, interest, and anxiety can be distinguished. (e) *Replication beyond the current base.* Multi-site studies in under-represented contexts, and replication of the three-profile trajectory solution in independent cohorts, would establish whether the patterns synthesised here generalise beyond the settings that produced them. Pre-registration and the routine reporting of sample sizes, effect sizes, and confidence intervals would materially improve the cumulative value of this literature, which at present reports these inconsistently.

Conclusion

This chapter has argued that the empirical literature on AI in foreign language education, read as a whole rather than study by study, is consistent with a single underlying mechanism: sustained, proficiency-calibrated feedback embedded within an iterative, documented process appears to move learners simultaneously along linguistic, cognitive, and affective lines of development in a way that isolated interventions do not. The pattern recurs across studies of writing, reading, speaking, and emotional self-regulation, and independently developed pedagogical frameworks arrive at compatible prescriptions from different starting points. That convergence is a stronger basis for a general claim than any single study offers, but—as the What Convergence Can and Cannot Establish subsection argued—it is not equivalent to a demonstrated causal mechanism, and the chapter has specified the designs that would settle the question.

From this, it is clear that the field is transitioning from assessment as a series of individual events to assessment as an ongoing and longitudinal process, with portfolios being the most logical existing option for realizing this transition, provided they are designed in a way that is iteration- and calibration-oriented and not as a storage system for completed work, which is where portfolio-based assessment has failed in practice (Abdulhamid & Fox, 2020). However, this transition comes with real limitations that should not be overlooked by the industry; especially, pragmatic competence does not go well with automated feedback for reasons that appear to be structural, rather than transitional (Liu et al., 2026); data governance becomes stricter as portfolios gather the records of activities for longer periods; and the amount of evidence supporting the process suffers

from being geographically, -relatedly, and bibliometrically limited (Andringa & Godfroid, 2020). If we consider these limitations, this body of work however appears to suggest a future in which foreign language teaching may respond more closely to the actual shape of language learning for each learner and people who teach languages will instead of being displaced will find more opportunity to use their judgment to explore these aspects of language competence that will stay human for a long time.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the author used ChatGPT and Grammarly in summer 2026 in order to improve grammar and language editing. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

Abdulhamid, N., & Fox, J. (2020). Portfolio based language assessment (PBLA) in Language Instruction for Newcomers to Canada (LINC) programs: Taking stock of teachers' experience. *Canadian Journal of Applied Linguistics*, *23*(2), 168–192. https://doi.org/10.37213/cjal.2020.31121

Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, *40*, 134–142. https://doi.org/10.1017/S0267190520000033

Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—A scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, *4*, Article 5. https://doi.org/10.1186/s41073-019-0064-8

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Ding, L., Zou, D., Kohnke, L., & Xie, H. (2025). Harnessing generative AI for L2 written corrective feedback: A systematic review of current research and future direction. In Proceedings of the 2025 2nd International Conference on Artificial Intelligence and Teacher Education (ICAITE) (pp. 1–6). IEEE. https://doi.org/10.1109/icaite68636.2025.11442331

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.2307/327317

Huang, W., Hew, K. F., & Fryer, L. K. (2021). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Huang, X., Wu, Y., & Dou, A. (2024). AI-enhanced task-based language teaching: Fostering personalized college English learning. *Frontiers in Educational Research*, *7*(10), 204–209. https://doi.org/10.25236/fer.2024.071033

Jeon, E. (2025). Artificial intelligence in ESL/EFL education: Evidence from recent reviews (2024–2025). *International Journal of Learning, Teaching and Educational Research*, *24*(10), 509–526. https://doi.org/10.26803/ijlter.24.10.24

Karakaya, K., Alpat, M. F., Uçar, H., Karakaya, Ö., & Bozkurt, A. (2025). Preparing teachers for the algorithmic educational landscape: A critical mapping of generative AI integration in language teacher education. *Technology in Language Teaching & Learning*, *7*(2), Article 102841. https://doi.org/10.29140/tltl.v7n2.102841

Khasawneh, M. A. S., Aladini, A., Assi, S. A., & Ajanil, B. (2025). Portfolio assessment in AI-enhanced learning environments: A pathway to emotion regulation, mindfulness, and language learning attitudes. *Language Testing in Asia*, *15*(1), Article 5. https://doi.org/10.1186/s40468-025-00345-0

Le, N. A. K., Tran, P. B. Q., & Tran, V. D. T. (2026). An investigation into the relationship between frequent AI use and English proficiency in portfolio-based learning among non-English majors. *HUFLIT Journal of Science*, *10*(2), 99. https://hjs.huflit.edu.vn/index.php/hjs/article/view/355

Li, B., Tan, Y. L., Wang, C., & Lowell, V. L. (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

Liu, M., Hashim, H., & Sulaiman, N. A. (2026). Longitudinal development of L2 speaking accuracy in AI-enhanced smart class: Evidence from an error-annotated learner corpus. *System*, *140*, Article 104048. https://doi.org/10.1016/j.system.2026.104048

Luo, Y., Xia, Y., & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*, *2*(2), 302–325. https://doi.org/10.1515/dsll-2025-0001

Nouzri, S., Fatimi, M. E., Guerin, T., Othmane, M., & Najjar, A. (2024). Beyond chatbots: Enhancing Luxembourgish language learning through multi-agent systems and large language models. In R. Arisaka, V. Sanchez-Anguix, S. Stein, R. Aydoğan, L. van der Torre, & T. Ito (Eds.), *Lecture Notes in Computer Science*: PRIMA 2024: Principles and Practice of Multi-Agent Systems (Vol. 15395, pp. 385–401). Springer. https://doi.org/10.1007/978-3-031-77367-9_29

Osawa, K. (2023). Integrating automated written corrective feedback into e-portfolios for second language writing: Notion and Notion AI. *RELC Journal*, *55*(3), 881–887. https://doi.org/10.1177/00336882231198913

Özçelik, N. P., & Ekşi, G. Y. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—A case study. *Smart Learning Environments*, *11*(1), Article 10. https://doi.org/10.1186/s40561-024-00296-8

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Rahman, A., Raj, A., Tomy, P., & Hameed, M. S. (2024). A comprehensive bibliometric and content analysis of artificial intelligence in language learning: Tracing between the years 2017 and 2023. *Artificial Intelligence Review*, *57*(4), Article 107. https://doi.org/10.1007/s10462-023-10643-9

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(1), Article 54. https://doi.org/10.1186/s41239-021-00292-9

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Uyanıker, P. (2017). Language assessment: Now and then. Avrasya Dil Eğitimi ve Araştırmaları Dergisi, 1(1), 1–20. https://dergipark.org.tr/en/pub/adea/article/287491

Vardi, I. (2011). The impact of iterative writing and feedback on the characteristics of tertiary students' written texts. *Teaching in Higher Education*, *17*(2), 167–179. https://doi.org/10.1080/13562517.2011.611865

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, C., Pek, L. S., Yuchen, Z., Latif, I. N. A., Zhou, B., & Li, M. (2026). Integrating generative AI in language education: A systematic review of pedagogical, ethical, and technological themes. *International Journal of Learning, Teaching and Educational Research*, *25*(2), 577–595. https://doi.org/10.26803/ijlter.25.2.26

Wang, R., Yin, Y., & Cao, Y. (2025). An LLM-based pedagogical framework for active, inquiry-based and adaptive learning in L2 writing. *Computers and Education: Artificial Intelligence*, *10*, Article 100535. https://doi.org/10.1016/j.caeai.2025.100535

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Xiao, Q.-M., & Liu, N.-H. (2025). The impact of AI-enhanced e-portfolios on EFL learners' writing proficiency and self-efficacy: A quasi-experimental study. In Proceedings of the 2025 9th International Conference on Education and E-Learning (pp. 245–250). Association for Computing Machinery. https://doi.org/10.1145/3789859.3789895

Yuan, H. (2025). Artificial intelligence in language learning: Biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities and Social Sciences Communications*, *12*(1), Article 556. https://doi.org/10.1057/s41599-025-04878-w

Zhang, S., & Liu, X. (2025). Learner emotions in AI-assisted English as a second/foreign language learning: A systematic review of empirical studies. *Frontiers in Psychology*, *16*, Article 1652806. https://doi.org/10.3389/fpsyg.2025.1652806

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zou, D. (2025). Leveraging artificial intelligence for enhanced language teaching and learning in higher education. In W. W. K. Ma, S. S. K. Cheung, C. Li, P. Prayadsab, & A. Mungwattana (Eds.), *Lecture Notes in Computer Science*: Blended Learning: Sustainable and Flexible Smart Learning (Vol. 15721, pp. 50–59). Springer. https://doi.org/10.1007/978-981-96-8430-4_4

Ethical Dimensions of Artificial Intelligence in Foreign Language Education

Zeyneb Dolan [1]

Abstract
The rapid integration of artificial intelligence (AI) and generative language models into foreign language education has transformed traditional teaching, learning, and assessment methods. While these technologies offer unprecedented opportunities for personalized learning, resource accessibility, and pedagogical efficiency, they simultaneously introduce profound ethical and pedagogical dilemmas. This paper examines the multi-faceted ethical dimensions of AI-enhanced language education, focusing on critical concerns such as student privacy violations and data security, algorithmic and linguistic biases, academic integrity threats like plagiarism, the risk of dehumanizing human-to-human interaction, digital divide inequalities, and opacity in algorithmic decision-making. Furthermore, it reviews prominent institutional governance frameworks and ethical guidelines established by international coalitions, including UNESCO, the European Commission, and the World Economic Forum, to operationalize transparency, equity, and human-centered oversight. Ultimately, the study argues that navigating this technological paradigm shift requires a proactive, principles-based framework combined with targeted digital literacy training to ensure that AI tools augment rather than replace human educators, safeguarding educational equity and human rights.
Keywords: Artificial Intelligence in Education, Foreign Language Education, Ethical Dimensions, Generative AI, Educational Equity

Introduction

Artificial Intelligence (AI) has transformed language teaching and learning by personalizing activities for individual learners, evaluating progress, and providing constant automated feedback (Akinsemolu & Onyeaka, 2025). As generative artificial intelligence advances, it steadily alters traditional models and methods of language education (Mohamed, 2024, p. 1). The integration of these technologies into language classrooms introduces a dynamic realm with profound ethical and pedagogical dimensions (Alghamdy, 2023). For instance, while platforms like ChatGPT have unlocked innovative avenues for language acquisition, they simultaneously introduce critical ethical dilemmas that demand thorough examination to guarantee conscientious deployment (Vaccino-Salvadore, 2023).

[1] Lecturer, Marmara University, School of Foreign Languages, English Prep School, ORCID: 0000-0001-6929-3161, *zeyneb.dolan@marmara.edu.tr*, Türkiye

Therefore, the instructional consequences of integrating AI into language classrooms necessitate a thorough investigation (Aljabr &Al-Ahdal, 2024).

Despite the evident benefits of AI, the handling and potential misuse of information used to train these systems, combined with algorithmic vulnerabilities, can expose users to completely unexpected hazards (Eden, Chisom & Adeniyi, 2024, Borenstein & Howard, 2020). On a larger scale, this requires separating the concept of doing ethical things from doing things ethically. Accordingly, educators must choose teaching methods that are fundamentally ethical while always accounting for the reality of unintended side effects (Holmes et al., 2021). Consequently, this technological shift highlights the urgency for policymakers and educators to examine the ethical implications of their practices, ensuring these tools benefit all students rather than reinforcing existing biases or inequalities (Compagnoni, Oguilve & Wen, 2025).

Crucially, although the extensive collection of user data such as voice recordings and interaction logs enables the personalized experiences that boost student engagement, it simultaneously creates serious vulnerabilities regarding data privacy, algorithmic transparency, and informed consent (Selvam & Vallejo, 2025). Therefore, while AI offers unprecedented opportunities for resource accessibility and pedagogical efficiency, it also imports systemic biases and compromises learner privacy (Chen, Chen & Li, 2025). Ultimately, ethical concerns in language learning arise when student autonomy is taken for granted to the extent that learners are assigned sole responsibility for their outcomes, regardless of their socioeconomic or educational backgrounds (Compagnoni, Oguilve & Wen, 2025).

Ethical Dimensions of Use of AI in Language Education

The widespread adoption of AI in language education raises significant ethical concerns regarding teacher job security and the potential for professional displacement due to automation (Alghamdy, 2023). Ethical implications also extend to the risk of deskilling educators, potentially devaluing the uniquely human facets of pedagogy such as empathy and cultural mentorship (Sabudu, 2025). Furthermore, the substantial computational resources required by large-scale AI models introduce a layer of ethical consideration concerning ecological sustainability and the environmental footprint of educational technology (Sabudu, 2025).

Critical questions also arise regarding intellectual property rights and the potential for copyright infringement when AI systems are trained on

datasets containing original work without explicit authorization (Sabudu, 2025; Pack & Maloney, 2024). Additionally, over-reliance on Western-centric datasets risks propagating a cultural monoculture that erodes global linguistic and cultural diversity (Chen, Chen & Li, 2025, Vaccino-Salvadore, 2023). Specifically in K-12 settings, unauthorized data collection and the inability of minors to provide truly informed consent present grave ethical and legal dilemmas (Selvam & Vallejo, 2025). To address these multi-faceted challenges, it is essential to examine a range of pressing concerns, including student privacy, algorithmic bias, plagiarism, limited human interaction, technology access inequalities, and systemic transparency issues.

Student Privacy and Data

Within the discourse on AI-enhanced language education, violations of student privacy represent one of the most critical ethical concerns (Mohamed, 2024). Significant security risks arise because AI-driven language learning platforms continuously gather and store extensive user data, such as writing samples, voice recordings, and individual learning patterns (Selvam & Vallejo, 2025). Because users often lack a clear understanding of what specific information is gathered and how far this collection goes, serious apprehensions exist regarding the potential exploitation of students' personal data (Alghamdy, 2023). Furthermore, the broad scale of data collection needed to power AI tools creates risks of unauthorized access, which can ultimately damage the trust between educators and their students (Mariyono et al., 2024). In response to these vulnerabilities, and to guarantee that learner data is handled ethically and protected from commercial exploitation, institutions must implement strong data protection frameworks and transparent consent procedures (Chen, Chen & Li, 2025).

Algorithmic Bias

In addition to privacy concerns, algorithmic bias poses a significant threat to educational

equity and can potentially perpetuate or amplify existing inequalities (Mariyono et al., 2024). Because AI algorithms reflect the values and perspectives of their developers, training datasets generally reflect historical and systemic biases (Mohamed, 2024). Consequently, large language models may exhibit persistent biases against gender, race, and religion, subtly shaping a user's opinions and behavior (Vaccino-Salvadore, 2023). Within language education, specifically, linguistic bias can occur when standard accents are prioritized, penalizing learners who have acquired

valid regional variations of a language (Chen, Chen & Li, 2025). This devalues the rich tapestry of the language-speaking world and reinforces a form of digital linguistic imperialism (Chen, Chen & Li, 2025). Therefore, bias mitigation is highly significant to prevent the unfair treatment of learners based on demographic attributes such as certain accents or socioeconomic status (Mohamed, 2024).

Plagiarism

Beyond systemic biases, the operational integrity of academic environments is also at risk. The widespread availability of AI tools has notably expanded the opportunities for students to engage in academic deceit such as plagiarism (Alghamdy, 2023). Teachers predominantly view the lack of original ideas and the submission of AI-generated text as one's own work as core forms of academic dishonesty (Cong-Lem, Tren & Nguyen, 2024). Moreover, over-reliance on chatbots like ChatGPT is considered detrimental to the development of building real comprehension, critical thinking capacities, and language proficiency (Cong-Lem, Tren & Nguyen, 2024). Students who leverage AI platforms as shortcuts to academic success could inadvertently bypass the cognitive involvement that underpins genuine learning (Alghamdy, 2023). Compounding this issue, large language models also have a tendency to hallucinate or fabricate information, and students may be unaware that the information they receive is inaccurate (Pack & Maloney, 2024).

Limiting Interactions

Alongside these academic integrity issues, the integration of AI in education introduces a risk of dehumanizing the learning process by reducing vital interactions between teachers and students (Mariyono et al., 2024). Human educators offer the empathy, emotional understanding, and rapport that serve as cornerstones for nurturing meaningful language learning experiences (Alghamdy, 2023). This shortfall in AI is evident in its complete incapacity to grasp the nuances of emotional intelligence or understand the complex web of learners' emotions (Alghamdy, 2023). Therefore, the guiding philosophy must dictate that AI functions as a tool to augment, rather than replace, the human educator (Chen, Chen & Li, 2025). Adopting a hybrid approach enables AI to manage repetitive drills and provide instant grammar feedback, thereby allowing teachers to focus their energy on nuanced cultural mentorship and building supportive classroom communities (Chen, Chen & Li, 2025).

Inequalities in Access to Technology

Simultaneously, inequalities in access to technology, particularly due to the digital divide, threaten to exacerbate existing socio-economic differences in language education (Akinsemolu & Onyeaka, 2025). Equitable access to AI-driven language learning technologies is significant for all learners regardless of their demographic or literacy level (Mohamed, 2024). However, inclusivity remains a challenge, as AI risks widening the educational gap between different socio-economic groups (Mariyono et al., 2024).

Furthermore, AI systems trained predominantly on Western datasets risk reinforcing existing inequities when deployed in diverse global contexts (Compagnoni, Oguilve & Wen, 2025). Therefore, addressing algorithmic oppression is essential to ensure that the tech industry does not marginalize underrepresented linguistic and cultural clusters (Mohamed, 2024).

Transparency Issues

Algorithmic opacity in AI systems severely hinders the effective implementation of core ethical principles (Niemi et al., 2023). Finally, these layered inequalities are compounded by a fundamental lack of clarity in how these systems operate. Because many AI models operate as black boxes, it is difficult for both educators and learners to comprehend the reasoning behind the assessments or recommendations they generate (Selvam & Vallejo, 2025). This absence of transparency creates an ethical obstacle by hindering metacognitive learning, given that students might be assigned a grade without understanding the underlying linguistic logic (Chen, Chen & Li, 2025).

Furthermore, accountability issues emerge when determining who is responsible for the biased feedback or errors that AI systems produce (Selvam & Vallejo, 2025). The decentralized structure of AI technologies spreads out responsibility, which ultimately complicates a student's ability to seek recourse for unfair algorithmic determinations (Chen, Chen & Li, 2025). Consequently, developers and educational institutions must implement explicit accountability frameworks alongside human oversight processes (Selvam & Vallejo, 2025).

Institutional Frameworks and Ethical Guidelines

To navigate these complex ethical dilemmas, a growing number of international governance bodies, educational organizations, and policy coalitions have established formal frameworks to steer the responsible

deployment of emerging technologies. As of 2021, more than 30 countries have instituted national AI policies designed to address the social, ethical, and structural impacts of artificial intelligence on key sectors like education (Schiff, 2021).

Prominent examples of these global guidelines include specific frameworks championed by the United Nations Educational, Scientific and Cultural Organization (UNESCO), the European Commission, and the World Economic Forum (WEF), each aiming to operationalize accountability, transparency, and equity on a systemic scale.

UNESCO's Ethical Framework for Generative AI in Education and Research

UNESCO's comprehensive guidance on generative AI (GenAI) in education and research (2023) is fundamentally anchored in a human-centered approach designed to ensure that technology serves the development of human capabilities for inclusive, just, and sustainable futures. This approach is guided by human rights principles, the protection of human dignity, and the preservation of cultural and linguistic diversity. In this framework, AI must not be allowed to usurp human intelligence; instead, it is viewed as an invitation to reconsider established understandings of knowledge and human learning.

To reassert public control and leverage the potential of GenAI across sectors, UNESCO proposes seven key regulatory steps for governmental agencies to regulate these tools. For data protection, governments should endorse international or regional general data protection regulations or develop national ones to regulate the collection and processing of personal data by GenAI providers. These efforts must be integrated into whole-of-government strategies, meaning that regulating GenAI must be integrated into broader national AI strategies that ensure safe and equitable use through coordinated intersectoral actions. Furthermore, countries must solidify and implement specific ethics regulations on the ethics of AI, translating guiding principles into enforceable laws. Copyright law adjustments are also required; existing laws should be adjusted or enforced to regulate AI-generated content, specifically regarding the use of copyrighted materials for training and the legal status of GenAI outputs. Additionally, national and local agencies need to develop GenAI-specific frameworks, including requirements for labeling AI-generated content. This must be paired with capacity building, where educational institutions, teachers, and researchers require support to strengthen their capacities for the proper, ethical, and effective use of GenAI. Finally,

governments should engage in long-term reflection, conducting open public debates and policy dialogues to examine the long-term implications of GenAI for knowledge creation, transmission, and validation.

In addition to governmental action, UNESCO outlines ten categories of accountabilities regarding ethical obligations for generative AI providers and developers to ensure their ethical deployment in educational contexts. Under human accountability, providers are responsible for adhering to core values, respecting intellectual property, and preventing the spread of disinformation and hate speech. They must also deliver trustworthy data and models by evidencing the trustworthiness of their data sources, ensuring legal compliance, and obtaining explicit consent for the use of personal information. Providers are strictly bound to provide non-discriminatory content, meaning the design and deployment of systems that generate biased or discriminatory content based on race, nationality, or gender must be prohibited. For explainability and transparency, providers should be transparent about their data sources, algorithms, and models' propensity for errors, offering explanations to public governance agencies. Regarding labeling and security, AI-generated content, such as papers, reports, images, and videos, must be clearly labeled as machine-produced, and providers must ensure secure and sustainable service throughout the system's life cycle. For safety and remediation, providers need to establish mechanisms for complaints and remedies, prevent predictable harms, and assist users in understanding the limitations of the generated content.

At the institutional level, governance and age restrictions must be strictly managed. Educational institutions hold the responsibility for determining which GenAI tools to procure and deploy within their environments. Institutional auditing mechanisms should be implemented to monitor algorithms, protect user data, and filter out inappropriate content. Crucially, UNESCO recommends a minimum age limit of 13 for the independent use of GenAI chatbots, and encourages countries to consider raising this threshold to 16 to protect children's rights and well-being. Institutional strategies must also address academic integrity by rethinking the design of written assignments. While GenAI providers are required to use watermarks, the long-term solution involves shifting assessment toward tasks that GenAI cannot perform better than humans, such as applying values like compassion and creativity to real-world challenges.

To prevent the widening of the digital divide and the risk of "data poverty," policy measures for inclusion and diversity must prioritize equity. To guarantee inclusive access, governments should provide universal connectivity and digital competencies, establishing funding for AI-enabled

tools for learners with disabilities or special needs. To protect linguistic and cultural pluralism, specifications should require AI providers to include data in multiple languages, particularly indigenous languages, and prevent the removal of minority languages or the imposition of dominant cultural norms. Finally, a critical evaluation of bias is necessary; users must be aware that GenAI outputs often represent only the dominant worldview and may marginalize minority voices due to the nature of their training data.

Furthermore, specific ethical guidelines for research and teaching must be established. The use of GenAI in research and teaching should be co-designed by teachers, learners, and researchers rather than being imposed in a top-down manner. For research, while GenAI can assist in literature reviews and data exploration, researchers must be alert to hallucinations or fabricated information, such as non-existent citations. For teaching, GenAI should be used to enhance teachers' understanding of their subjects and methodologies, and to facilitate higher-order thinking through pedagogically appropriate interactions. To safeguard human agency, policies must prevent the use of GenAI where it would deprive learners of opportunities to develop cognitive abilities, social skills, and independent logical reasoning through real-world observations.

The rapid evolution of GenAI also presents several uncharted ethical issues and future risks that require immediate and ongoing investigation. First, the psychological and developmental impact must be monitored, as systems that mimic human interaction may have unknown effects on the cognitive development and emotional well-being of young learners. There is also the threat of a reduction of human connection, as the use of AI systems in education may reduce the critical social-emotional aspects of learning that occur through human-to-human interaction. Additionally, an over-reliance on GenAI for suggestions and outlines may lead to a homogenization of thought, causing a standardization of responses and weakening the value of independent inquiry and diverse creative output. Finally, users must look out for hidden discrimination, since future iterations of GenAI may generate new forms of bias and discrimination that are difficult to detect due to the non-transparent "black box" nature of artificial neural networks.

The European Commission's Ethical Guidelines

The European Commission's 2026 revised guidelines (2026) emphasize that artificial intelligence (AI) is fundamentally reshaping education, necessitating a human-centered approach to ensure technology serves people and supports educators. Integrating an ethical dimension into

educational AI is not merely an addition but a central component of digital and AI literacy. These guidelines are grounded in the belief that educational applications of AI must respect human dignity, promote fairness, and preserve the integrity of learning.

The primary legal framework governing AI use in the European Union is the Artificial Intelligence Act (2024), which adopts a risk-based approach to ensure systems are safe and transparent. Under this Act, certain AI practices are strictly prohibited in educational settings, most notably the use of AI systems designed to detect or infer a person's emotions, unless for specific medical or safety reasons. Furthermore, AI tools used for sensitive tasks such as admissions, grading, and student progress tracking are classified as high-risk due to their potential impact on fundamental rights. High-risk systems are subject to strict obligations, including adequate risk mitigation, high-quality datasets to prevent discriminatory outcomes, and human oversight. In addition to the AI Act, the General Data Protection Regulation (GDPR) applies to all processing of personal data in schools, requiring institutions to act as data controllers who communicate clearly about data usage. Educational institutions are required to carry out Data Protection Impact Assessments (DPIA) and, for high-risk tools, Fundamental Rights Impact Assessments (FRIA) to identify and mitigate risks to individual rights. Transparency is a key legal requirement, as users have a "right to explanation" to understand the logic behind AI-assisted decisions that significantly affect them.

To underpin the use of AI in teaching and learning, the Commission identifies five foundational ethical pillars that must be upheld across all educational environments:

- Human Dignity: This encompasses the right to privacy and human autonomy, ensuring individuals are treated as intrinsically valuable rather than mere data objects.
- Fairness: This principle requires that all users have equal access to opportunities through inclusion, non-discrimination, and the equitable distribution of rights.
- Trust/Trustworthiness: AI is considered trustworthy when it is consistently reliable, respects privacy, avoids bias, and promotes the best interests of students.
- Academic Integrity: This extends beyond tools to include mindsets and competencies, focusing on using AI in an honest manner and designing valid assessments in the era of generative AI.
- Justified Choice: This requires that collective choices in the school environment are based on knowledge and participatory models of decision-making.

Building upon these foundational pillars, the guidelines establish eight key operational requirements for the ethical deployment of AI, structured through guiding questions for teachers and school leaders:

- Human Agency and Oversight: Teachers must be able to review and adjust AI outputs before they are used in teaching and intervene if inappropriate effects are observed.
- Transparency and Explainability: Providers must supply clear documentation on how a tool works, its decision logic, and its treatment of intellectual property.
- Diversity and Inclusion: AI-assisted materials must adequately represent different languages, cultures, and abilities, and schools must address digital-literacy gaps that prevent equal usage.
- Fairness and Non-Discrimination: Systems must be tested for bias across different learner groups to prevent misrepresentation or exclusion of marginalized populations.
- Societal and Environmental Wellbeing: Educators should monitor whether AI use affects student motivation, mood, or social interaction, while also considering the environmental impact of the technology.
- Privacy and Data Governance: Sensitive student data must be anonymized, and educators should avoid inputting personally identifiable information into external AI tools.
- Technical Robustness and Safety: Schools must establish formal procedures for reporting technical faults, unsafe outputs, and data breaches.
- Accountability: There must be a clear division of responsibilities for support, maintenance, and the evaluation of whether AI tools align with educational values.

In the specific context of language learning, AI tools offer specialized support such as real-time translation, speech recognition, and interactive conversation simulations that adapt to a learner's proficiency level. These tools can provide automated feedback on pronunciation, fluency, and comprehension to enhance the learning process. However, when using generative AI for language tasks, educators must encourage students to verify outputs, as models may produce "hallucinations"—coherent but fabricated or factually incorrect information. Discussing these hallucinations is a practical way to develop AI literacy, helping students understand that AI generates text based on statistical patterns rather than true understanding.

World Economic Forum Education 4.0

The World Economic Forum (2024) defines Education 4.0 as a comprehensive framework for teaching and learning that focuses on providing learners with the abilities, skills, attitudes, and values fit for a future defined by accelerating technological change. While artificial intelligence (AI) offers a unique opportunity to enhance the reach and effectiveness of these education systems, its deployment necessitates careful management to navigate the associated risks.

A fundamental ethical principle in the integration of AI is that technology must serve to enhance, not replace, the role of the teacher. While AI can excel at streamlining administrative duties and presenting differentiated content, the complex process of facilitating learning remains an indispensable human element. By automating routine clerical tasks which currently consume a significant portion of educators' time AI empowers teachers to focus on high-value human interactions, such as building relationships, providing social-emotional support, and offering individualized mentorship. This synergy is intended to improve teaching effectiveness while preserving the social-emotional aspects of learning that occur through human connection.

The integration of AI systems raises significant ethical concerns regarding the safeguarding of student data and individual privacy. Stakeholders, including parents and educators, have expressed ambivalence or a lack of trust in AI systems due to fears of exploitation by bad actors. To address these concerns, the World Economic Forum emphasizes that robust data privacy and security protocols must be implemented at the institutional and governmental levels. These protocols should include:

- Raising awareness about the ethical use of data.
- Ensuring explicit consent and anonymizing sensitive student information.
- Limiting data collection to what is strictly necessary for educational purposes.
- Developing governance measures and equity frameworks before the rapid deployment of generative AI.

Ethical education in the age of AI requires the cultivation of AI and digital literacy, which extends beyond the mere technical ability to use tools. This literacy encompasses critical thinking, problem-solving, and an awareness of the ethical implications of AI design and deployment. Students must be equipped with the discernment to:

- Assess the reliability of sources and the accuracy of information provided by AI.

- Recognize and critically analyze disinformation and misinformation, particularly as synthetic content becomes more prevalent.
- Understand the limitations, reliability, and potential unconscious biases inherent in machine learning algorithms.
- Navigate the ethical considerations of being a "responsible producer" of technology, which includes understanding information privacy, security, and the potential societal impacts of AI design.

A primary ethical risk identified by the World Economic Forum is the potential for AI to exacerbate existing education gaps and widen global inequality. There is a significant concern that if the benefits of AI primarily accrue to privileged communities and geographies, the education equity gap will further increase. Therefore, equity and inclusion considerations must be central to the design of all AI-integrated educational programs. Ethical implementation requires:

- Universal connectivity and efforts to reach the 2.6 billion individuals who currently lack basic internet access.
- The use of Universal Design for Learning (UDL) principles to create inclusive tools for diverse learners, including the 240 million children globally who have disabilities.
- Leveraging AI to provide customizable interfaces and adaptive technologies, such as real-time captioning for students with auditory impairments or text-to-speech functions.
- The development of AI models that address language barriers by providing training and support in diverse local and indigenous languages.

To balance innovation with ethical guardrails, the World Economic Forum proposes five essential conditions for the integration of AI in education. First, AI tools must be developed in active collaboration with educators and education leaders to ensure they meet diverse learning needs. Second, sensitive information must be protected through rigorous security protocols. Third, innovative funding models are required to support independent testing and evaluation of AI solutions to ensure their efficacy and value. Fourth, teachers, students, and administrators must receive the necessary training and upskilling to make productive and ethical use of AI systems. Finally, a focus on academic integrity is necessary, as the rise of generative AI challenges basic assumptions about how classroom work reflects a student's cognitive process and necessitates new ways to discourage cheating.

In the specific realm of language education, generative AI presents unique ethical and pedagogical challenges by mimicking human logic, writing,

and discourse. These advancements put into question the relevance of traditional methods for teaching writing, grammar, and logic. However, AI-driven tools also offer transformative potential for literacy development by providing real-time constructive feedback and personalized learning journeys. For language learners, AI can:

- Provide tailored content and adjust the pace of instruction based on individual proficiency levels.
- Offer instant translation to adapt culturally relevant materials to a student's lived experience and background.
- Enable interactive and engaging learning through gamified assessments and virtual mentors that provide 24/7 support.

Conclusion

The integration of Artificial Intelligence into language education marks a profound paradigm shift, redefining the boundaries of pedagogical efficiency, personalization, and student engagement. However, as this paper has demonstrated, these technological advancements are intrinsically bound to complex ethical dilemmas that cannot be ignored. Over the course of this analysis, several pressing vulnerabilities have been identified, most notably regarding the exploitation of student privacy through massive data harvesting and the perpetuation of systemic inequities via algorithmic and linguistic biases. Furthermore, the widespread availability of generative tools poses severe threats to academic integrity by facilitating plagiarism, risks dehumanizing the classroom by reducing vital teacher-student interactions, and threatens to widen the global socio-economic divide due to unequal technology access. Finally, the "black box" nature of many modern algorithms complicates these issues further, obscuring transparency and dividing accountability when systems produce errors or biased outcomes.

To mitigate these multifaceted challenges, the global educational community must shift away from localized, reactive troubleshooting and move toward systemic, policy-driven solutions. As observed in the regulatory frameworks put forward by prominent international governing bodies, managing the deployment of AI requires coordinated, structural oversight. Landmark guidelines from UNESCO, the European Commission, and the World Economic Forum collectively underscore that technology must remain human-centered, serving to augment rather than replace human pedagogy. By enforcing strong data protection standards, mandating algorithmic risk assessments, prohibiting discriminatory or invasive AI practices, and addressing the global digital divide, these international initiatives

offer foundational blueprints for balancing technical innovation with human rights and educational equity.

Moving forward, resolving the tension between technological capability and ethical responsibility requires a concerted effort from educators, developers, and policymakers alike. To foster a more informed and ethical approach, universal standard language pedagogy should be used to develop learning paths and assessments for AI applications (Akinsemolu & Onyeaka, 2025). It is essential that training is provided to both learners and educators to develop the necessary digital literacy skills to use and critically evaluate AI tools effectively (Compagnoni, Oguilve & Wen, 2025). Rather than viewing ethics as a restrictive barrier, ethical awareness functions as a cognitive prerequisite for active evaluation, rather than merely inhibiting technology use (Zhang & Li, 2026). Empowering students in this manner yields practical behavioral benefits; for instance, students with a holistic ethical mindset are significantly more likely to engage in critical verification and reflection when using AI translation tools (Zhang & Li, 2026). Ultimately, a proactive, principles-based ethical framework is a prerequisite for the responsible and sustainable integration of AI into the language education ecosystem (Chen, Chen & Li, 2025).

Declaration on the Use of Generative Artificial Intelligence

Throughout the creation of this manuscript, the author employed ChatGPT (OpenAI, GPT-5) solely to assist with textual refinement and translation editing. All theoretical foundations, data interpretations, and academic conclusions were independently conceived, reviewed, and finalized by the author, who maintains complete accountability for the academic integrity and ethical standards of the work.

References

Akinsemolu A. A., & OnyeakaH. (2025). The Role of Artificial Intelligence in Transforming Language Learning: Opportunities and Ethical Considerations. *Journal of Language and Education*, *11*(1), 148-152. https://doi.org/10.17323/jle.2025.22118

Alghamdy, R. Z. (2023). Pedagogical and ethical implications of Artificial intelligence in EFL context: A review study. *English Language Teaching*, *16*(10), 1-87. doi: 10.5539/elt.v16n10p87

Aljabr, F. S., & Al-Ahdal, A. a. M. H. (2024). Ethical and pedagogical implications of AI in language education: An empirical study at Ha'il University. *Acta Psychologica*, *251*, 104605. https://doi.org/10.1016/j.actpsy.2024.104605

Borenstein, J., & Howard, A. (2020). Emerging challenges in AI and the need for AI ethics education. *AI And Ethics*, *1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Chen, J., Chen, J., & Li, Y. (2025). Analysis of ethical dilemmas and avoidance strategies in the application of artificial intelligence in international Chinese language education. *Frontiers in Sustainable Development*, *5*(9), 163–172. https://doi.org/10.54691/1zwqnx79

Compagnoni, I., Oguilve, V., & Wen, W. (2025). Ethical Frontiers of AI in Language Pedagogy: Navigating transparency, accountability, and equity. *International Journal of Linguistics*, *17*(4), 1. https://doi.org/10.5296/ijl.v17i4.22768

Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative ai: perceptions and responses of vietnamese efl teachers. *Teaching English with Technology, 24* (1), 28–47. https://doi.org/10.56297/FSYB3031/MXNB7567

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, *10*(2), 006–013. https://doi.org/10.30574/msarr.2024.10.2.0039

European Commission. (2026). *Guidelines on the ethical use of artificial intelligence and data in teaching and learning for educators* (Revised ed.). Luxembourg: Publications Office of the European Union.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). Ethics of AI in Education: Towards a Community-Wide framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Li, J., & Huang, Q. (2025). Navigating ethical Dilemmas in AI-Enhanced education. *International Journal of Knowledge Management*, *21*(1), 1–21. DOI: 10.4018/IJKM.382384

Mariyono, D., Hidayatullah, A. N. A., & Kamila, A. N. A. (2024). Navigating Ethical Dilemmas in AI-Enhanced Language Education: Addressing bias and ensuring Inclusivity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4936530

Mohamed, M. S. P. (2024). Exploring ethical dimensions of AI-enhanced language education: A literature perspective. *Technology in Language Teaching & Learning*, *6*(3), 1813. https://doi.org/10.29140/tltl.v6n3.1813

Pack, A., & Maloney, J. (2024). Using Artificial Intelligence in TESOL: Some ethical and pedagogical considerations. *TESOL Quarterly*, *58*(2), 1007–1018. https://doi.org/10.1002/tesq.3320

Sabudu, D. (2026). Ethical Aspects of AI in Language Learning Management: A Literature review. *Enrichment Journal of Multidisciplinary Research and Development*, *3*(9), 3709–3720. https://doi.org/10.55324/enrichment.v3i9.579

Selvam, M., & Vallejo, R. G. (2025). Ethical and privacy considerations in AI-Driven Language learning. *LatIA*, *3*, 328. https://doi.org/10.62486/latia2025328

Schiff, D. (2021). Education for AI, not AI for Education: The Role of Education and Ethics in National AI Policy Strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527–563. https://doi.org/10.1007/s40593-021-00270-2

United Nations Educational, Scientific and Cultural Organization. (2023). *Guidance for generative AI in education and research*. Paris, France: UNESCO Publishing.

Vaccino-Salvadore, S. (2023). Exploring the ethical dimensions of using ChatGPT in language learning and beyond. *Languages*, *8*(3), 191. https://doi.org/10.3390/languages8030191

Niemi, H., Pea, R. D., & Lu, Y. (2022). *AI in Learning: Designing the Future*. Springer Nature.

World Economic Forum. (2024). *Shaping the future of learning: The role of AI in education 4.0* (Insight Report). Geneva, Switzerland: World Economic Forum.

Zhang, L., & Li, Q. (2026). Ethical perceptions and learning engagement in AI-assisted translation: a behavioral psychology study of English learners. *Frontiers in Psychology*, *17*, 1739478. https://doi.org/10.3389/fpsyg.2026.1739478

About the Series

Contemporary Studies on Digital Humanities, edited by Ali Tilbe, Yusuf Topaloğlu and Gökhan Dinar, is an international, peer-reviewed book series devoted to the study of human expression, knowledge and culture in an age when the tools of reading, writing, teaching and interpreting are increasingly computational. Published by Maurer Press (Frankfurt am Main) in cooperation with AYG Press İstanbul, the series brings together scholars from literary studies, linguistics, language education, translation studies, history, philosophy, cultural and media studies, the arts and the information sciences. It welcomes research on computational approaches to texts and cultures, digital archives and heritage, corpus-based inquiry, artificial intelligence in scholarship, teaching and creative practice, digital pedagogy and assessment, and the ethical and epistemological questions these developments raise. Approaching technology neither as a miracle nor as a threat, but as a new condition of humanistic work, the series seeks to build a plural and rigorous dialogue on the digital transformation of the humanities, with particular attention to research from Türkiye and its neighbouring academic communities alongside contributions from across the globe. It publishes monographs, edited volumes and thematic collections.

About the Editors

Ali Tilbe (Prof. Dr.), Professor of French Language Education and Head of Department at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education. His research focuses on contemporary and Francophone literature, autofiction, sociocriticism and the teaching of French as a foreign language.

Yusuf Topaloğlu (Assoc. Prof. Dr.), Associate Professor of French Language Education at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education. His research focuses on French linguistics and syntax, written production in French as a foreign language and artificial intelligence in language education.

Gökhan Dinar (Asst. Prof. Dr.), Assistant Professor of French Language Education at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education, and the department's Erasmus coordinator. His research focuses on postcolonial Francophone literature, sociocriticism, artificial intelligence in language education and translation studies.

Previously Published

Vol. I:	From Generative to Agentic Artificial Intelligence in Foreign Language Teaching, Learning, and Assessment *by Tuncer Can*
Vol. II:	Between Code and Canon Artificial Intelligence in Language and Literature Studies *by Gökhan Dinar & Yusuf Topaloğlu*
Vol. III:	New Directions in Foreign Language Education Pedagogies for a Changing World *by Yunus Emre Sarı & Ali Tilbe*