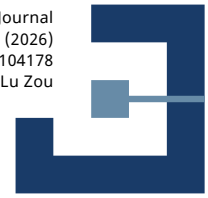
The JALT CALL Journal
ISSN 1832-4215, 22, 104178 (2026)
https://doi.org/10.29140/jaltcall.2026.104178
©2026 Lu Zou



Castledown



This work is licensed under a Creative Commons Attribution 4.0 International License.

AI-enhanced blended learning for German as a second foreign language: An embedded case study in higher education

Lu Zou
Beijing International Studies University, China
zzxp111@126.com

The integration of AI into blended language education has renewed discussion on how AI-supported analytics and feedback can be pedagogically incorporated into classroom practice. This embedded case study examines an AI-enhanced blended course in higher education, where German was taught as a second foreign language at the first-semester level. It explores how AI-supported analytics – including Automatic Speech Recognition (ASR) pronunciation scoring, AI-generated error reports, and AI-supported feedback – and teacher-created online resources were integrated within a teacher-mediated instructional design. The study combined platform-generated analytics, item-level assessment records, structured teacher reflection logs, and a post-semester student questionnaire. Findings indicate that these analytics increased the visibility of learner engagement patterns and recurrent linguistic difficulties – particularly in German pronunciation and grammatical structures requiring multiple rule applications – enabling targeted instructional adaptation. Their pedagogical value, however, depended on the teacher's professional interpretation and contextual judgement; systematic discrepancies between AI-generated outputs and teacher evaluation confirmed that AI-supported feedback complemented rather than replaced teacher-led pedagogy. Learner agency emerged primarily through autonomy-supportive engagement within teacher-designed tasks – including selective review, pacing decisions, and strategic use of error logs – rather than through independent AI exploration. These findings suggest that AI-enhanced blended learning is pedagogically valuable when analytics function within a teacher-mediated instructional system, where visibility informs instructional adaptation through professional judgement and learner engagement emerges through pedagogically structured participation, reinforcing a human-centred approach to AI-enhanced language education within a human–AI shared regulation perspective.

Keywords: AI-enhanced blended learning; AI-supported analytics; teacher mediation; learner agency; German as a second foreign language

Introduction

AI, CALL, and German as a foreign language

Artificial Intelligence (AI) has become an increasingly prominent component of second language (L2) education, shaping feedback, practice, assessment, and learner engagement (Godwin-Jones, 2021; Huang et al., 2021; Kasneci et al., 2023). While early applications focused on discrete functions such as pronunciation training through Automatic Speech Recognition (ASR) and scripted conversational practice (Dizon & Tang, 2020), generative AI and large language models have expanded AI-supported learning towards more flexible interaction, automated feedback, and learning analytics (Godwin-Jones, 2024). AI has moved from experimentation towards routine application, becoming a normalised component of contemporary language learning environments, with recent studies documenting its pedagogical role as a scaffolding resource within specific language learning contexts (Huang & Cassany, 2025; Kasneci et al., 2023). For Warschauer and Xu (2024), this development is transformative: generative AI is reshaping how second language learners access information, produce text, and participate in academic practices – with stakes "especially high for second language learners" (p. 2) – making it "much more than a pedagogical tool" (p. 2). Yet whether such transformation fulfils its promise depends on pedagogical integration – a question that CALL research and German as a Foreign Language (GFL) scholarship have long addressed.

Research has demonstrated positive effects of AI-assisted learning on vocabulary, listening, speaking, and writing, as well as reduced language-learning anxiety (Wiboolyasarin et al., 2025), yet recent reviews reveal substantial variation across skills, learner populations, and instructional designs. Learning outcomes are shaped primarily by pedagogical context and implementation; meta-analyses confirm that learning environment, feedback configuration, teacher comprehensibility, and intervention duration significantly moderate AI effectiveness (Wu, 2024; Lee & Lee, 2024). Research on blended language learning converges on the same pattern: outcomes depend largely on how online and face-to-face components are structured, sequenced, and integrated within coherent instructional designs (Jehoul et al., 2025). Across both domains, this variation underscores a longstanding principle: how technology is framed and mediated matters more than whether it is used.

These observations resonate with a longstanding CALL tradition emphasising pedagogical integration over technological innovation – a principle extensively elaborated within GFL pedagogy (Rösler, 2010; Rösler & Würffel, 2013; Baum & Günay, 2025), the field known in German as Deutsch als Fremdsprache (DaF). Within GFL scholarship, researchers have consistently argued that digital tools derive educational value from their alignment with instructional goals, learning activities, and teacher guidance rather than from technological

sophistication itself (Rösler, 2008, 2010; Rösler & Würffel, 2010, 2013). Although contemporary AI technologies introduce new possibilities through automated feedback, learning analytics, and generative language support, they do not fundamentally alter this principle: technology acquires educational significance through instructional design and teacher mediation rather than through functionality alone (Alm, 2024; Kohnke et al., 2023; Biebighäuser & Feick, 2020; Schwab & Hoffmann, 2024).

The rapid expansion of generative AI has nevertheless intensified debates concerning automation, teacher roles, learner autonomy, and the relationship between human and machine contributions to language education (Zawacki-Richter et al., 2019; Luckin & Cukurova, 2019; Warschauer & Xu, 2024). In response, this study adopts a classroom-based, teacher-mediated perspective, conceptualising AI-supported tools not as autonomous agents but as pedagogically framed resources whose educational value emerges through professional judgement, instructional design, and situated classroom practice.

Research gap and study objectives

Despite the substantial body of research on CALL and digital media in German language education (Rösler, 2008, 2010; Rösler & Würffel, 2010, 2013), scholarly discussion within GFL pedagogy has largely focused on learner autonomy, online interaction, and digitally mediated communication within broader CALL frameworks (Rösler, 2014, 2016, 2018; Biebighäuser & Feick, 2020; Schwab & Hoffmann, 2024), leaving the pedagogical integration of AI-supported analytics, generative AI tools, and automated feedback systems within teacher-mediated GFL learning environments largely unexplored. This gap is compounded by German's distinctive linguistic demands for beginning learners – rich inflectional morphology encompassing verb conjugation and noun/adjective declension, verb-second and verb-final word order patterns, and phonological contrasts such as umlaut vowels (ö and ü) – suggesting that findings from English-dominated AI research may not transfer directly to GFL contexts.

This gap is not unique to GFL: it mirrors a broader trend in CALL research. While studies have examined individual AI technologies (Dizon & Tang, 2020; Huang et al., 2021), less is known about their pedagogical integration within authentic classroom settings – a need identified by Warschauer and Xu (2024). Stockwell and Reinders (2019) caution against assuming that technological access fosters meaningful learner autonomy, and Guan et al. (2024) confirm that while AI tools may promote self-regulation, they do not necessarily foster motivation or meaningful autonomy. These observations direct attention to how pedagogical mediation shapes learner engagement with AI-supported resources – a question particularly salient in GFL contexts, where the longstanding emphasis on teacher-mediated learning (Rösler, 2010; Rösler & Würffel, 2013) makes the relationship between AI capabilities and instructional design especially consequential.

This study investigates AI integration in a teacher-led blended German as a second foreign language course in higher education. Focusing on a first-semester

course, it examines how AI-supported analytics, the teacher's professional judgement, and instructional mediation interacted to shape learner engagement, the exercise of learner agency, and the development of autonomy. To address these questions, the study draws on a pedagogical mediation model linking instructional visibility, pedagogical mediation, and learner agency; the model is developed in the following section (see Figure 1) and is operationalised in the Design Rationale and Data Analysis sections of the Methodology.

Theoretical framework

Teacher mediation in AI-enhanced blended learning

Blended learning integrates face-to-face instruction with technology-mediated activity to extend opportunities for practice, reflection, and interaction – while maintaining teacher guidance (Garrison & Vaughan, 2008; Graham, 2013). Its pedagogical value lies in the deliberate coordination of online and offline components; effectiveness depends on how these components are sequenced, aligned, and integrated within instructional design (Boelens et al., 2017). Pedagogical orchestration – teachers' ongoing regulation of tasks, tools, and timing to support engagement (Dillenbourg, 2013) – has long been documented in technology-mediated language learning, where teacher presence is enacted through task design and material selection rather than direct interaction (Lai et al., 2015; Reinders & Benson, 2017).

Teacher judgement becomes even more salient in AI-enhanced environments. While AI increases the visibility of learning processes, analytic outputs acquire pedagogical relevance only through interpretation – made in relation to curricular goals, learner needs, and classroom contingencies. This relationship is increasingly framed through hybrid intelligence, emphasising complementarity rather than substitution (Jarrahi et al., 2022; Cukurova, 2024): AI contributes pattern detection and data aggregation; teachers remain responsible for contextual interpretation, pedagogical reasoning, and determining how data should inform scaffolding, feedback, and task adjustment.

Professional judgement develops through deliberate engagement rather than spontaneously from tool availability. Liu et al. (2024) documented teachers' progression from initial reliance on AI to more critical and coordinated use, while Liu and Xiao (2024) found that engagement with generative AI remained concentrated at material preparation rather than extending into curriculum-embedded practice – confirming that judgement must be deliberately channelled towards classroom integration.

Learning analytics are promoted as tools for enhancing instructional awareness – making patterns of engagement and performance visible that might otherwise go unnoticed (Clow, 2012; Bergdahl et al., 2020). Visibility alone, however, does not change learning outcomes: data must be interpreted and acted upon (Clow, 2012). Without that second step – interpretation and action – visibility remains merely informational rather than pedagogically transformative. Analytics are frequently underutilised or restricted to monitoring, risking

surveillance framing (Slade & Prinsloo, 2013; Prinsloo, 2017; Ley et al., 2023). These limitations extend to AI-supported conversational interaction: human–AI dialogue systems show systematic differences in openings and closings compared with human–human interaction, indicating that current AI dialogue tools cannot replicate authentic communicative exchange (Dombi et al., 2024).

Accordingly, this study approaches AI-enhanced blended learning through an orchestration-oriented lens. Pedagogy, not technology, provides the primary analytic framework.

Learner engagement and autonomy

Learner autonomy – the capacity to take responsibility for one's own learning (Holec, 1981; Little, 2007) – is not innate but develops through scaffolded participation and mediated engagement grounded in sociocultural principles (Lantolf & Thorne, 2006), particularly within instructional contexts that provide structured opportunities for decision-making and reflection (Benson, 2011; Little, 2007).

Within CALL, Blin (2004) called for richer theoretical tools to examine technology–autonomy relationships, while Grgurović et al. (2013) demonstrated comparable effectiveness – though comparable outcomes do not imply technology functions independently of instructional design. Recent discussions similarly stress that AI-supported learning requires pedagogical guidance (Kohnke et al., 2023; Warschauer & Xu, 2024).

Despite this evidence, much discourse infers autonomy from independent tool use – an assumption overlooking the cognitive and strategic demands of effective learning (Stockwell & Reinders, 2019). This study therefore adopts a mediated view: autonomy as a pedagogically grounded outcome developing through scaffolded participation, and agency as emerging through interaction with instructional structures rather than from independent tool access alone.

In the present study, learner autonomy is conceptualised as a developmental capacity that emerges through scaffolded participation, while learner agency refers to the situated exercise of choice within structured learning environments. Agency constitutes the observable manifestation of emerging autonomy – decisions about how to engage with resources, feedback, and tasks – and can be exercised within teacher-guided settings. The operationalisation of learner agency through autonomy-supportive engagement indicators is described in the Methodology section.

Pedagogical mediation model

Synthesising the preceding discussions, this study proposes an analytics-informed pedagogical mediation model. The model conceptualises AI-enhanced blended language learning as a mediated instructional system organised around four interrelated dimensions: AI-supported analytics, teacher professional judgement, instructional mediation, and learner engagement and agency. AI functions as augmentative infrastructure – it enhances instructional

visibility and extends feedback capacity, but it does not determine pedagogical action. Decision-making remains anchored in the teacher's professional judgement, which integrates analytic information with curricular goals, classroom interaction, and contextual constraints (Stockwell & Reinders, 2019). Figure 1 summarises the conceptual relationships among these dimensions and serves as a guiding framework for the empirical analysis; the operationalisation of this model for data collection and analysis is presented in the Methodology section.

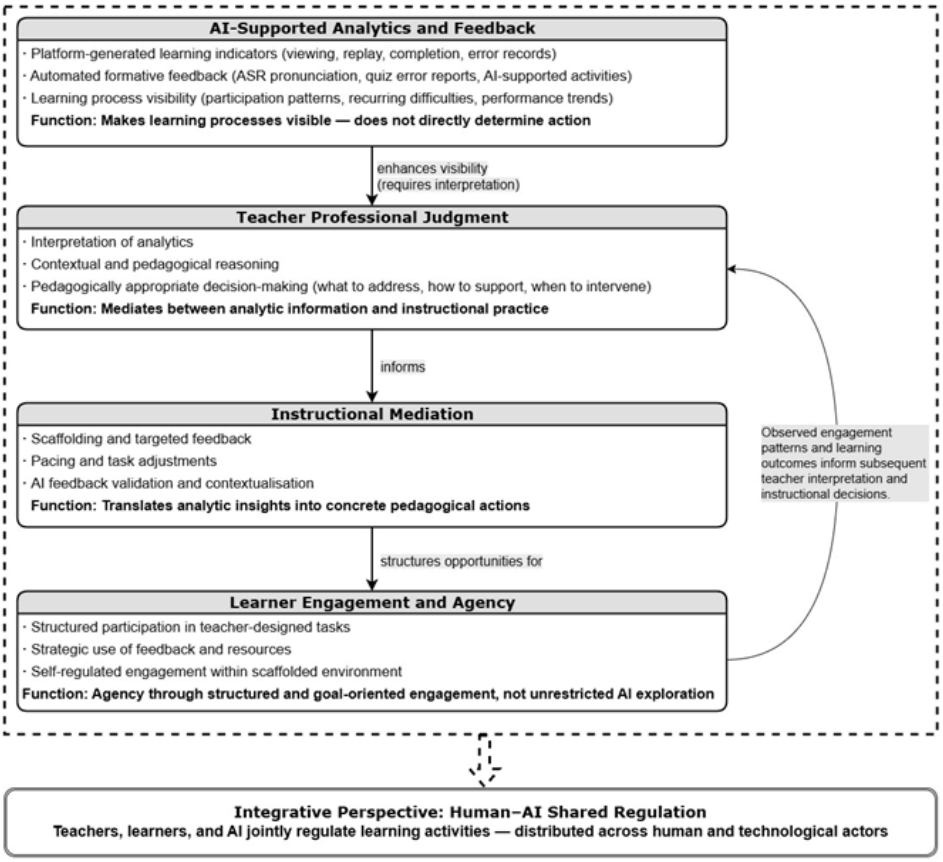


Figure 1. Pedagogical mediation model of AI-enhanced blended learning for German as a second foreign language

Research context

Course context

The study was conducted in a first-semester blended course at a foreign language university in China, where German was taught as a second foreign language following students' primary foreign language, which included English, French, Japanese, Russian, Italian, and Hungarian. Participants were 11 undergraduate students (4 female, 7 male) from diverse disciplinary backgrounds.

The first semester was selected because pedagogical mediation is particularly

important at this stage, where students are simultaneously developing foundational linguistic knowledge, learning strategies, and participation routines. Examining AI-enhanced blended learning at this juncture makes visible the processes of scaffolding and adjustment that later become tacit.

The course was designed and taught by the instructor-researcher, combining online preparation with face-to-face instruction. The online environment comprised 83 instructional sections with 245 resources (52 released micro-videos, 193 pronunciation tasks) and 877 quiz items embedded across tasks. Teacher-created micro-videos addressing phonetics, grammar, vocabulary, and situational communication formed the core online resources, aligned with A1–B2 curriculum progression; each video followed a consistent internal sequence (explanation, read-along practice, exercises, walkthrough).

Although designed primarily for pre-class preparation, platform analytics showed that many students revisited videos after classroom sessions, suggesting dual pedagogical functions within the blended cycle. Rather than encouraging extensive use of external AI tools, the course emphasised teacher-curated resources supported by platform-based AI functionalities. This was not a technical limitation but a design choice: for most participants, German was a second foreign language, requiring substantial cognitive investment within limited instructional time. The instructional approach foregrounded teacher-led design, maintaining instructional coherence while using AI-generated information to enhance awareness of learner progress.

Instructional orchestration

The instructional design integrated online and face-to-face components through deliberate teacher orchestration. Online activities primarily supported preparation, consolidation, and self-review, while classroom sessions focused on clarification, guided practice, and communicative application.

The online component was delivered through the SuperStar Learning Platform (Chaoxing Fanya; https://www.chaoxing.com/), a widely used learning management system in Chinese higher education. The platform integrates instructional delivery, learning analytics, and AI-supported feedback within a single cloud-based environment that supports content hosting (videos, documents, quizzes), automated assessment, and data collection. AI-supported functions – including automated error categorisation and generative feedback – are powered by an integrated DeepSeek large language model, which processes learner inputs and generates corrective suggestions within the platform's task interface. Crucially, the platform does not determine instructional sequencing, learning pathways, or assessment decisions; these remain under teacher control.

Grammar and vocabulary activities provided rule-based feedback, while pronunciation tasks employed ASR technology to generate scores for pronunciation accuracy, fluency, and completeness. Automated pronunciation scores were routinely reviewed by the teacher and manually adjusted where the automated evaluation was judged inaccurate; the pronunciation scores reported

in this study reflect these teacher-verified ratings. Students had access to individualised error logs that categorised recurring errors, supporting self-monitoring and review. Platform-generated analytics captured engagement indicators including video viewing duration, task completion rates, replay behaviour, item-level response records, and quiz performance.

Face-to-face sessions served as the primary site of pedagogical mediation. Classroom activities included grammar consolidation, guided dialogues, translation exercises, oral communication tasks, and cultural discussions. Analytics-informed insights were used diagnostically to identify areas requiring additional support, review, or clarification. Decisions concerning pacing, scaffolding, task design, and instructional adjustment remained under teacher control and were informed by the combined consideration of analytics, classroom interaction, and professional judgement.

Research questions

Guided by the theoretical framework of pedagogical orchestration, mediated learner agency, and analytics-informed pedagogical mediation, this study addresses the following research questions:

RQ1. How did students engage with AI-supported activities and feedback within a teacher-mediated blended German as a second foreign language course, as reflected in engagement patterns, task participation, and reported learning practices?

RQ2. How did the teacher interpret and utilise AI-supported analytics to inform instructional decision-making and pedagogical adjustment over the course duration?

RQ3. How did students exercise learner agency through autonomy-supportive engagement with the teacher-mediated AI-enhanced blended learning environment?

To clarify the analytical focus of each question, RQ1 addresses engagement patterns and performance (what students did and how they performed); RQ2 addresses teacher interpretation and use of AI-supported analytics (how the teacher interpreted the analytics and translated them into instructional adaptation); and RQ3 addresses learner agency (how students exercised choice within the teacher-designed structures). The three questions are complementary, and the alignment between research questions, data sources, and analytical procedures is summarised in Table 1.

Table 1. Mapping of research questions, data sources, and analytical procedures

RQ	Data sources	Analytical procedures	Reported in
RQ1	Platform-generated analytics; item-level assessment records (SM3); structured teacher reflection logs (SM2); post-semester student questionnaire (SM1)	Descriptive statistics at class and individual level; summary of open-ended responses	Results RQ1 (Tables 3–6, see also Table 8 under RQ3)
RQ2	Structured teacher reflection logs (SM2)	Deductive thematic analysis using the coding framework (Table 2)	Results RQ2 (Table 7)
RQ3	Post-semester student questionnaire (SM1); structured teacher reflection logs (SM2); platform-generated analytics	Thematic coding of open-ended responses (Table 2); descriptive statistics; cross-checked against reflection logs and platform analytics	Results RQ3 (Table 8; SM1, Sections 3.2–3.3)

Note. SM = supplementary material (SM1–SM4); full titles are listed in the Data Availability Statement. Data sources are those specified in the Data Sources section. RQ1 was addressed through descriptive indicators and was not coded qualitatively; the coding framework applied to RQ2 and RQ3 is presented in Table 2. For RQ2, the data sources listed are those coded; platform-generated analytics and item-level assessment records (SM3) provided additional evidence in the Results.

Methodology

Design rationale

This study adopted an embedded case study design with a qualitatively driven mixed-methods orientation to investigate how AI-supported analytics were pedagogically used within a blended German as a second foreign language course, focusing on a single instructional setting. Rather than evaluating effectiveness, the study sought to understand how AI-generated information was interpreted and enacted through teacher mediation. Four complementary sources of evidence were collected (see Data Sources below), following the principle of integrating multiple forms of evidence for comprehensive understanding (Creswell & Plano Clark, 2017). Quantitative indicators served a descriptive function – documenting engagement patterns and recurrent difficulties to contextualise qualitative interpretation – rather than being treated as measures of effectiveness; qualitative evidence played the central role in examining how analytics-informed insights were translated into pedagogical decisions.

The pedagogical mediation model (see Figure 1) structures the analysis around four operational functions derived from its conceptual dimensions: visibility via AI-supported analytics, interpretation via teacher professional judgement, adaptation via instructional mediation, and structured participation via learner engagement and agency. Observed engagement patterns subsequently inform teacher interpretation, creating a feedback loop within the model.

Data sources

The study drew on four complementary sources of evidence: (1) platform-generated analytics, (2) item-level assessment records, (3) structured teacher reflection logs (including classroom observation records), and (4) post-semester student questionnaire.

Platform-generated data comprised system analytics (viewing duration, replay behaviour, completion rates, pronunciation scores, AI-generated error reports) and item-level assessment records from quiz and pronunciation data across seven instructional units, identifying recurrent error patterns as descriptive evidence for pedagogical interpretation.

Classroom processes were documented through structured teacher reflection logs. To enhance consistency across instructional units, a standardised template was used throughout the semester, comprising four sections: observation, diagnosis, instructional response, and outcome evaluation. Reflection records focused on learner participation, comprehension, translation performance, use of situational expressions, and responsiveness to feedback. Building on this continuous observation of classroom performance, the teacher additionally rated each student's in-class situational dialogue performance holistically (0–100) for each of the six unit themes; these ratings are reported in Table 5. Together, the reflection logs and dialogue ratings enabled the triangulation of platform-generated indicators with observed classroom behaviour and instructional decision-making.

At the end of the semester, students completed a confidential, pseudonymised questionnaire capturing perceptions of the AI-enhanced blended learning environment. The questionnaire comprised 29 Likert-scale items across six dimensions (overall learning experience, AI-supported functions, teacher-created online resources, teacher mediation support, learner agency, perceived learning outcomes; see Table 8 for structure) and three open-ended questions. All Likert items used a five-point scale (1="strongly disagree" to 5="strongly agree"). Given the small sample size, internal consistency (Cronbach's Alpha) was not computed; results are treated as indicative perceptions rather than psychometrically validated measures. Responses were linked to platform data through pseudonymised codes (Student A–K) for triangulation; personally identifiable information was removed prior to analysis. The full questionnaire is provided in SM1.

Data analysis

Quantitative and qualitative data were analysed in relation to the three research questions. Given the small sample size (N=11), quantitative analysis served a descriptive rather than inferential purpose; all identified issues were pedagogically validated through teacher reflection logs. Qualitative data from teacher reflection logs and open-ended questionnaire responses were analysed thematically: iterative reading and coding identified recurring patterns; codes were then grouped into categories and refined into themes. This analysis was

conducted by the instructor-researcher (the sole author), who also designed and taught the course; the full coding manual is provided in SM4.

For RQ1, analysis addressed three dimensions: engagement patterns, task participation, and reported learning practices. Descriptive indicators (completion rates, viewing duration, replay ratios, quiz scores, pronunciation and oral task performance) were compared at class and individual levels. Reported practices were examined through questionnaire responses alongside teacher reflection logs.

For RQ2, teacher reflection logs were analysed by the instructor-researcher using deductive thematic analysis. Entries containing references to platform data or instructional decisions were identified across the seven instructional units, deductively coded using the framework in Table 2, and grouped into three themes: interpreting analytics as indicators of learning difficulty; translating analytics into instructional adaptation; and validating and contextualising AI-generated feedback through professional judgement.

For RQ3, learner agency was operationalised through observable autonomy-supportive engagement (selective replay, pacing decisions, targeted use of feedback) rather than as a fixed psychological construct. The instructor-researcher examined the qualitative data – open-ended questionnaire responses and teacher reflection logs – to capture resource-supported learning strategies and autonomy-supportive engagement; these were then compared with platform-generated indicators for consistency.

Findings were triangulated across all four data sources; AI-supported analytics were treated as contextual indicators requiring pedagogical interpretation rather than objective measures. Table 2 presents the coding framework applied to the qualitative data, mapping refined codes, categories, and themes to RQ2 and RQ3. Grounded in the pedagogical mediation model, the framework addresses instructional visibility and pedagogical mediation through the RQ2 themes, and learner agency through the RQ3 dimensions.

Table 2. Qualitative coding framework

RQ	Refined codes	Category	Theme
RQ2	score discrepancies, recurring errors, replay concentration, completion–performance mismatch	analytics interpretation	Interpreting analytics as indicators of learning difficulty
RQ2	targeted review, activity redesign, pacing adjustment, teacher modelling, feedback clarification	instructional adaptation	Translating analytics into instructional adaptation
RQ2	AI score verification, error interpretation, triangulation with classroom performance	teacher mediation	Validating and contextualising AI-generated feedback through professional judgement
RQ3	replaying videos, reviewing error logs, selective revision	learning strategies	Resource-supported learning strategies
RQ3	independent task completion, feedback use, learning management	autonomy-supportive engagement	Autonomy-supportive engagement as structured and scaffolded agency

Note. RQ1 was addressed through descriptive indicators rather than qualitative coding and is therefore not listed. The two RQ3 rows present analytical dimensions rather than themes. The categories and themes are data-driven refinements of the pedagogical mediation model's functions rather than one-to-one operationalisations of its dimensions. Full code definitions, inclusion and exclusion criteria, illustrative examples, and decision rules are provided in SM4.

Researcher positionality and reflexivity

The researcher occupied multiple roles as course designer, instructor, and analyst. This combination of roles provided direct access to instructional processes but introduced the possibility of interpretive bias; the observed centrality of teacher mediation may partly reflect that the researcher designed the instructional framework and interpreted the resulting data.

To reduce this risk, several measures were implemented: classroom observations were documented immediately after each lesson using a structured reflection template (observation, diagnosis, instructional response, outcome evaluation); interpretations were triangulated across platform analytics, assessment records, and questionnaire responses; and analytics were treated as provisional indicators requiring pedagogical validation.

The findings should therefore be understood as contextually situated interpretations of classroom processes rather than objective evaluations independent of the researcher's instructional role.

Results

Learner engagement and learning outcomes (RQ1)

To address RQ1, findings are reported across three dimensions consistent with the analytical framework described in the Methodology: (1) engagement patterns (platform-generated viewing, replay, and completion indicators); (2) task participation (quiz scores, oral task performance, and course achievement);

and (3) reported learning practices (questionnaire responses describing how students used AI-supported resources).

Course achievement was indexed by the platform-generated composite score (0–100), composed of two equally weighted components: (a) an online task completion rate (maximum 50 points), covering completion of video lessons and their accompanying exercises, pronunciation tasks and their accompanying exercises, and the independent unit tests, irrespective of the scores obtained; and (b) the average score on the assigned section quizzes (maximum 50 points), which comprised the exercises accompanying the video lessons, the ASR-scored pronunciation tasks, and the independent unit tests (students could attempt each item twice, with their higher score counting; uncompleted items counted as zero). Section quizzes thus entered both components – through completion in (a) and through scores in (b).

Across the 11 participants, the semester mean composite score was 90.82 out of 100 (SD=24.97). This variance was concentrated in one outlier rather than distributed across the cohort: ten students' composite scores ranged from 96.80 to 99.53 (mean=98.34, SD=1.02), while Student K – who enrolled late – obtained a composite score of 15.58. Section quiz scores averaged 96.69 out of 100 across the ten completing students (89.29 across the full cohort of 11), and the average task completion rate (completed ÷ assigned tasks × 100) was 92.35%; excluding Student K, the remaining ten students achieved a completion rate of 100%. Student K completed only 15.86% of the assigned online tasks; pronunciation (89.14) and grammar (90.63) averages nonetheless remained within the cohort range, and the composite score of 15.58 corresponded closely to the 15.86% completion rate.

Engagement patterns, the first dimension, were examined through platform-generated viewing and replay indicators. As shown in Table 3, total viewing time ranged from 222.3 minutes (Student K) to 1091.4 minutes (Student D), with a cohort mean of 587.7 minutes. Weighted replay ratios varied from 95.2% (Student K) to 467.7% (Student D), with substantial variation across students. Students D, B, and G recorded the highest total viewing times and weighted replay ratios; Student K showed the lowest. The maximum single-video replay ratio peaked for Student C (3088.74%). Replay activity was highest for teacher-created videos addressing pronunciation, grammar, and language use.

Table 3. Platform-generated viewing and replay indicators by student (N=11)

Student	Total viewing time (min)	Weighted Replay Ratio (%)	Max replay ratio (%)
A	579.4	248.2	2213.79
B	776.0	332.7	1866.93
C	628.4	269.3	3088.74
D	1091.4	467.7	2825.28
E	436.8	187.2	1171.84
F	529.1	226.6	977.95
G	693.3	297.1	941.43
H	459.4	196.7	703.07
I	405.9	173.9	2033.03
J	642.2	275.2	1946.13
K	222.3	95.2	152.22

Note. The weighted replay ratio = a student's total video viewing time ÷ the combined original duration of all course videos, ×100; values above 100% indicate repeated viewing (e.g., 248.2% ≈ 2.5 times the full video corpus). The max replay ratio (%) = the highest per-video value of (cumulative viewing time on a single video ÷ that video's original duration, ×100); very high values (e.g., 3088.74%) typically reflect repeated looped playback of short pronunciation modelling clips for shadowing practice.

Task participation, the second dimension, was examined through quiz scores, oral task performance, and course achievement. Table 4 presents students' average scores on the section quizzes targeting the ten core grammatical structures of the semester syllabus. Scores were highest for present tense and possessive pronouns, whereas adjective declension and stative–dynamic prepositions produced the lowest scores across the cohort. Table 5 presents the teacher's ratings of students' in-class situational dialogue performance for the six unit themes; ratings were highest in earlier units and tended to decline as communicative demands increased. Table 6 presents individual course performance data; ten of eleven students achieved composite scores ranging from 96.80 to 99.53, while Student K obtained a composite score of 15.58 with a 15.86% task completion rate.

Table 4. Grammar performance by student and structure (10 grammar points)

Student	Present tense	Imperative	Possessive pronouns	Accusative	Dative	Stative–dynamic prepositions	Adjective declension	Person–object order	Modal verbs	Reflexive verbs
A	100	95	100	95	90	88	85	92	90	92
B	98	92	95	93	88	85	82	88	87	89
C	100	97	95	96	92	90	88	90	91	93
D	95	90	90	92	85	82	80	85	86	87
E	90	85	88	86	80	78	76	82	80	83
F	93	88	90	89	83	80	78	84	82	85
G	96	90	92	91	86	83	80	86	85	88
H	92	87	88	87	82	79	77	83	81	84
I	88	82	85	84	78	75	73	80	78	81
J	95	90	92	90	85	82	80	86	84	87
K	80	75	78	76	70	68	65	72	70	74
Mean	93	88	90	89	84	81	79	84	83	86

Note. Each cell reports the student's average score across all section-quiz items assessing the corresponding grammatical structure (0–100; platform-recorded; students could attempt each item twice, with their higher score counting). The ten structures constitute the core grammar syllabus of the first semester across the seven instructional units; for each column, the instructor-researcher identified all quiz items assessing that structure (e.g., items from the verb-conjugation, imperative, and possessive-pronoun quizzes) and averaged the student's recorded scores on these items. The final row (Mean) presents the cross-student average for each structure.

Table 5. Structured oral task performance (6 situational dialogues)

Student	Test 1 Introduction	Test 2 Campus	Test 3 Family	Test 4 Restaurant	Test 5 Living In Germany	Test 6 Shopping
A	95	92	94	90	88	91
B	93	90	92	88	86	89
C	98	96	97	95	94	96
D	90	88	89	85	83	86
E	82	78	80	75	73	76
F	85	82	84	80	78	81
G	88	85	87	83	81	84
H	84	80	82	78	76	79
I	80	75	78	72	70	74
J	90	88	89	85	83	86
K	75	70	72	68	65	69
Mean	87	84	86	82	80	83

Note. Each cell shows the teacher's holistic rating (0–100) of the student's in-class situational dialogue performance for the corresponding unit theme (Introduction, Campus, Family, Restaurant, Living in Germany, Shopping). Ratings were based on the teacher's observation of dialogue performance throughout each unit, considering task completion, accuracy, and fluency; they are teacher ratings, not platform-generated scores, and are reported separately from the composite score (Table 6).

Table 6. Student-level course performance and task completion

Student	Composite score	Task completion component (/50)	Pronunciation tasks completed / total	Video lessons completed / total	Section quiz score component (/50)	Pronunciation avg	Grammar avg
A	99.45	50 / 50	193 / 193	52 / 52	49.45 / 50	98.76	99.58
B	99.25	50 / 50	193 / 193	52 / 52	49.25 / 50	98.42	96.88
C	97.19	50 / 50	193 / 193	52 / 52	47.19 / 50	96.35	97.08
D	99.53	50 / 50	193 / 193	52 / 52	49.53 / 50	99.02	99.67
E	97.17	50 / 50	193 / 193	52 / 52	47.17 / 50	95.11	78.96
F	97.91	50 / 50	193 / 193	52 / 52	47.91 / 50	96.84	90.21
G	98.39	50 / 50	193 / 193	52 / 52	48.39 / 50	97.06	88.54
H	98.72	50 / 50	193 / 193	52 / 52	48.72 / 50	96.91	92.13
I	99.03	50 / 50	193 / 193	52 / 52	49.03 / 50	97.88	98.42
J	96.80	50 / 50	193 / 193	52 / 52	46.80 / 50	94.73	77.92
K	15.58	7.93 / 50	9 / 193	26 / 52	7.65 / 50	89.14	90.63

Note. Composite Score = task completion component (weighted 50%) + section quiz score component (weighted 50%), as generated by the platform. The Task Completion Component (/50) and Section Quiz Score Component (/50) columns report the two components. The task completion component was awarded for completion of the course's 454 assigned tasks – 52 released micro-videos, 193 pronunciation tasks, and the independent unit tests – irrespective of scores; the Pronunciation Tasks Completed and Video Lessons Completed columns present the completion counts for the pronunciation and video task subsets underlying this component. The section quiz score component was the average score on all assigned section quizzes, which comprised the exercises accompanying the video lessons, the ASR-scored pronunciation tasks, and the independent unit tests; uncompleted quizzes were scored as zero (students could attempt each item twice, with their higher score counting). Pronunciation Avg (mean ASR score of the 193 pronunciation tasks, teacher-verified – see Results, RQ2, Theme 3) and Grammar Avg (mean across grammar-related section quizzes) are both descriptive subsets of the section quiz score component; Grammar Avg spans the full range of grammatical content across the semester and is therefore not directly comparable with the ten-structure scores in Table 4.

Students' reported learning practices, the third dimension of RQ1, were captured through the post-semester questionnaire. In open-ended responses, the most frequently mentioned AI-supported resources were speech correction, automated error records, and automatic scoring with explanatory feedback. Students described using error logs to review weak points and to revisit knowledge points they did not understand. On Likert-scale items, AI-supported functions received a mean rating of 4.65 (SD=0.61) and teacher-created online resources 4.83 (SD=0.39). These descriptive findings address RQ1's focus on engagement patterns; the agency dimension of these practices is addressed in the RQ3 section.

In summary, RQ1 asked how students engaged with AI-supported activities and feedback. Three findings converge: (1) participation was near-universal, but engagement intensity varied substantially (viewing time 222–1091 min; weighted replay 95–468%), with replay concentrated on pronunciation and grammar videos – the domains where quiz data showed persistent difficulty; (2) performance was uniformly high (composite scores 96.80–99.53) except for the late-enrolled Student K, whose language scores nonetheless remained within the cohort range; (3) students reported strategic, diagnostic use of AI functions (speech correction, error logs, and automated scoring with explanatory

feedback) rather than undifferentiated task completion. Engagement in this environment was thus high but selective and differentiated.

Teacher interpretation and pedagogical use of analytics (RQ2)

Analysis of teacher reflection logs generated three interconnected themes describing how platform-generated information was interpreted and incorporated into instructional practice throughout the semester. Table 7 summarises the themes, categories, and representative codes derived from the thematic analysis; the coding categories were developed deductively from the research questions and refined inductively through iterative analysis of the reflection logs. Throughout this section, individual cases and log excerpts are labelled "illustrative case" or "illustrative excerpt"; recurring phenomena or instructional responses observed across multiple units are labelled "cross-unit pattern"; and unlabelled statements present each theme's overall coding results.

Table 7. Themes derived from teacher reflection logs (RQ2)

Theme	Category	Representative codes
Theme 1. Interpreting analytics as indicators of learning difficulty	analytics interpretation	score discrepancies, recurring errors, replay concentration, completion–performance mismatch
Theme 2. Translating analytics into instructional adaptation	instructional adaptation	targeted review, activity redesign, pacing adjustment, teacher modelling, feedback clarification
Theme 3. Validating and contextualising AI-generated feedback through professional judgement	teacher mediation	AI score verification, error interpretation, triangulation with classroom performance

Note. Themes and categories follow the coding framework in Table 2. Coding was applied to the structured teacher reflection logs (SM2). The evidence underlying each theme is identified in the text, where illustrative cases and excerpts are labelled with their data source and Lektion (German for "lesson"; the course's seven instructional units are numbered Lektion 1–7).

Theme 1: Interpreting analytics as indicators of learning difficulty. Analysis of platform-generated records, item-level assessment records, and teacher reflection logs identified recurring learning difficulties across pronunciation, grammar, and communicative transfer domains.

Illustrative case (Lektion 1: item-level assessment records and teacher reflection logs). Detailed examination of these records identified recurring pronunciation difficulties. Front-rounded vowels (short ö [œ] and long ö [ø:]) yielded accuracy rates between 72.7% and 81.8%, with L1 Mandarin speakers substituting back vowel equivalents. Final-obstruent devoicing proved particularly challenging, with accuracy rates as low as 27.3% (e.g., *Kind, Daten*); learners frequently produced voiced stops where the German standard requires devoicing. Vowel-length distinctions (e.g., *ob* [ɔ] vs. *Boot* [o:], *Mund* [ʊ] vs. *gut* [u:]) showed elevated error rates – the ASR system applied native-speaker timing criteria to beginner-level productions, penalising slower speech rates as fluency deficits – and context-sensitive *ch* alternation (*ich* [ɪç] vs. *Buch* [x])

yielded rates as low as 54.5%. Teacher logs documented systematic discrepancies between automated scores and manual evaluation, noting factors including poor microphone quality, background noise, native-speaker criteria applied to beginners, and penalisation of untaught phonemes.

Cross-unit pattern (subsequent units). Analysis identified recurring grammatical difficulties. In Lektion 4, items requiring co-occurrence of multiple grammatical demands showed higher error rates than isolated-rule items. In later units, persistent difficulties with stative–dynamic prepositions and adjective declension appeared; completion rates remained high while spoken hesitation continued.

Theme 2: Translating analytics into instructional adaptation. A second theme concerned how analytic insights were translated into instructional adaptation. As the difficulties documented in Theme 1 recurred across instructional units, reflection logs recorded targeted pedagogical responses informed by analytic interpretation.

Cross-unit pattern (pronunciation and oral-practice adaptation). In pronunciation instruction, teacher modelling and minimal-pair exercises followed the accuracy rates reported in Theme 1. Across several units, activities were redesigned to provide more controlled oral production with immediate teacher feedback when quiz performance showed rule recognition alongside spoken hesitation.

Illustrative case (grammar adaptation, Lektion 4). Errors clustered in items requiring simultaneous application of multiple rules. Illustrative excerpt (teacher reflection log, Lektion 4): “Several students could identify the correct verb form in isolation but produced errors when separable verb placement and case marking co-occurred.” Classroom activities were redesigned to decompose sentences into functional components before introducing grammatical analysis; for stative–dynamic prepositions, additional contrastive explanations, guided practice, and targeted review were introduced; and communicatively useful imperative forms were prioritised in oral practice.

Theme 3: Validating and contextualising AI-generated feedback through professional judgement. The third theme concerned teacher mediation in evaluating the pedagogical significance of AI-generated information. Reflection logs documented instances in which automated feedback was reviewed against classroom evidence before instructional action was taken.

Illustrative case (pronunciation assessment, teacher reflection logs, Lektion 1). Systematic discrepancies between automated scores and manual evaluation were documented: recordings assigned scores of 72.7 and 77.8 were judged as fully accurate after teacher review; a student with consistently excellent pronunciation received a system score of 83 while the teacher rated 100; and scores of 92.7 and 91.8 were manually upgraded to 100. These discrepancies corresponded to the factors documented in Theme 1 – poor microphone quality, background noise, native-speaker criteria applied to beginners, and penalisation of untaught phonemes.

Illustrative case (Lektion 5: validation of completion analytics against classroom performance). Later in the semester, platform data showed high completion rates and elevated replay activity, but item-level records revealed recurring errors in verb–preposition combinations and word order, with 6 of 20 comprehensive test items below 80% accuracy. Classroom observation corroborated this discrepancy: students demonstrated surface-level familiarity but hesitated during spontaneous interaction. The teacher interpreted this as evidence that task completion did not indicate communicative readiness; instruction in Lektion 5 was restructured around controlled oral practice with immediate teacher feedback.

Illustrative case (Lektion 6 and 7). In the final two units, automated feedback was explicitly repositioned. In Lektion 6, explicit reminders were issued that automated feedback should be treated as a reference rather than an authoritative judgement. In Lektion 7, AI feedback was positioned as a background support tool during integrated review activities. Across these instances, AI-generated information was treated as descriptive input within a teacher-guided instructional cycle.

In summary, RQ2 asked how the teacher interpreted and used AI-supported analytics. Across the three themes, a consistent mediation cycle emerged: analytics surfaced learning difficulties (Theme 1); these interpretations were translated into targeted instructional adaptations – task redesign, pacing adjustment, teacher modelling (Theme 2); and AI-generated feedback was continuously validated against classroom evidence before informing action (Theme 3). Analytics thus informed, but did not determine, pedagogical decision-making.

Autonomy-supportive engagement and learner agency (RQ3)

RQ3 examined how learners exercised agency within the AI-enhanced blended environment. Consistent with the pedagogical mediation model, learner agency was operationalised through autonomy-supportive engagement – selective, strategic choices about resource use, feedback engagement, and pacing within the teacher-designed instructional framework – rather than through task participation rates alone. Findings are organised around one theme with two analytical dimensions: resource-supported learning strategies and autonomy-supportive engagement.

As shown in Table 8, students reported positive perceptions across all dimensions. Teacher-created online resources received the highest rating (M=4.83, SD=0.39), followed by overall learning experience (M=4.80, SD=0.49); teacher mediation support (M=4.73, SD=0.65) and perceived learning outcomes (M=4.73, SD=0.65) were rated similarly. AI-supported functions were also evaluated positively (M=4.65, SD=0.61). Responses related to learner agency showed the greatest variation among the six dimensions (M=4.59, SD=0.99).

Table 8. Student perceptions questionnaire results

Dimension	Items	N	Min	Max	M	SD
Overall learning experience	A1–A4 (4)	11	3.50	5.00	4.80	0.49
AI-supported functions	B1–B5 (5)	11	3.40	5.00	4.65	0.61
Teacher-created online resources	C1–C7 (7)	11	3.71	5.00	4.83	0.39
Teacher mediation support	D1–D3 (3)	11	3.00	5.00	4.73	0.65
Learner agency	D4–D6+E1 (4)	11	1.75	5.00	4.59	0.99
Perceived learning outcomes	E2–E7 (6)	11	3.00	5.00	4.73	0.65

Note. N=11. Likert-scale items used a 5-point scale (1="strongly disagree" to 5="strongly agree"). Internal consistency was not computed; see the Methodology section for the descriptive orientation of questionnaire results. The full questionnaire is provided in SM1.

Individual responses varied in both what students chose to report and their overall evaluative stance: Student C – whose single-video replay ratio peaked at 3088.74% – explained that the platform "helps me review knowledge points I don't understand," while Student A juxtaposed appreciation of prompt error correction ("Errors can be corrected promptly") with the limitation that feedback was "not targeted enough." At the other end of the range, Student F, the only respondent scoring below 3 on agency items, and Student D, despite the cohort's highest engagement (weighted replay 467.7%), both offered no open-ended commentary; Student H valued automated scoring while noting the platform "sometimes freezes due to network issues." The dispersion of the agency dimension (SD=0.99) reflects these contrasting individual patterns rather than a uniformly moderate level across the cohort (see SM1, Section 3.3; SM2, Section 5).

Open-ended responses and platform analytics showed differentiated patterns of resource use. Students most frequently identified speech correction, automated error records, and AI-supported review functions as useful resources, helping them identify weaknesses and review difficult content. As Student B noted: "The AI teaching assistant has been very helpful; the error collection helps me review my weak points." Platform analytics corroborated these reports with specific strategy–resource alignments: high replay ratios were concentrated on pronunciation and grammar videos – the same domains where quiz data revealed persistent difficulties (see Table 4) – indicating that selective replay served as a self-initiated remediation strategy rather than unfocused repetition. Error logs, in turn, supported targeted review: students who referred to them described using the recorded errors to revisit weak points and monitor recurring mistakes, whereas others did not mention error logs in their open-ended responses (see SM1, Section 3.2).

Teacher reflection logs and platform analytics further documented variation in how students used available learning resources: some learners repeatedly revisited instructional videos and completed multiple review attempts, whereas others relied more heavily on classroom clarification and teacher explanation. Students also identified limitations in AI-supported feedback – including insufficient targeting, pronunciation recognition inaccuracies, and

occasional technical problems – and reported evaluating AI-generated feedback in relation to their individual learning needs rather than accepting it uncritically.

In the questionnaire, students who reported strategic, self-directed use of AI-supported resources (selective review, targeted feedback engagement, independent pacing decisions) tended to record higher agency scores. At the other end of the range, the lowest agency rating came from the student who most consistently sought face-to-face explanation: classroom observation recorded attentive listening, active participation, and frequent engagement with conceptual and cultural extensions of the curriculum, alongside an explicit preference for teacher explanation (see SM2, Section 5) – a pattern indicating that reduced reliance on AI-supported functions reflected a learning preference rather than disengagement. At the same time, the cohort as a whole rated teacher mediation support highly on average (M=4.73, SD=0.65; Table 8), suggesting that orientation towards teacher guidance was not confined to less independent learners. Both patterns constitute agency, exercised through different mediational pathways within the same teacher-designed framework. These patterns document variation in how students engaged with AI-supported resources, though not all individual response patterns were equally open to interpretation. The alignment between agency ratings and self-reported strategy use is a descriptive observation rather than a statistically verified relationship: with N=11, correlational analysis was precluded, and this alignment rests on questionnaire self-reports triangulated against platform indicators and teacher observation (SM2).

Discussion

This study examined how AI-supported analytics functioned within a teacher-mediated blended German as a second foreign language course. Rather than evaluating effectiveness, the study explored how AI-supported information was interpreted and enacted within classroom practice.

Learner engagement as differentiated and goal-oriented participation

The engagement data revealed consistently high levels of task completion and quiz performance, accompanied by substantial variation in viewing behaviour and replay activity. These differentiated patterns suggest goal-oriented resource selection: high replay ratios were concentrated on linguistically demanding topics such as pronunciation, adjective declension, and stative–dynamic prepositions, indicating that students selectively invested effort in areas they perceived as particularly challenging. AI-supported quizzes thus served a dual role – functioning as consolidation tools that supported practice and review while simultaneously revealing areas in which engagement did not immediately translate into mastery. Persistent difficulties with adjective declension when combined with other grammatical rules, and with stative–dynamic

prepositions, illustrate how AI-supported analytics can make underlying conceptual challenges visible despite generally high completion rates.

The persistence of conceptual difficulties despite high completion rates echoes previous research cautioning that engagement patterns should not be interpreted solely in quantitative terms: analytics metrics risk optimising for what is measured rather than what is valued (Clow, 2012), and learning analytics applications in blended settings, including the platform-generated indicators used in the present study, have been criticised for dispersed theoretical and methodological grounding (Bergdahl et al., 2020). Engagement, moreover, is often selective and varied, reflecting learners' perceived needs, goals, and learning preferences rather than uniform patterns of resource use (Lai et al., 2015). This caution extends to aggregate descriptive statistics: the ten completing students' composite-score standard deviation of only 1.02 on a 100-point scale – compared with an overall cohort SD of 24.97 inflated by a single case of non-completion – illustrates how cohort-level variability indicators can misrepresent the actual distribution of learning outcomes when one outlier disproportionately inflates them. Making these engagement patterns visible did not, however, automatically generate instructional responses; their educational significance depended on professional interpretation.

Teacher mediation as the pedagogical core

Teacher mediation functioned as the critical mechanism translating analytic information into meaningful instructional action – aligning with hybrid intelligence conceptualisations (Jarrahi et al., 2022; Cukurova, 2024), where AI contributes pattern detection while teachers retain responsibility for interpretation and adaptation. Two distinct domains required such professional contextualisation. In engagement analytics, the Lektion 5 findings (see Theme 3) demonstrated that completion metrics alone could not establish learning achievement; the teacher's interpretation of high completion rates as procedural rather than communicative readiness constituted a pedagogical judgement that analytics could not generate independently. In pronunciation assessment, systematic discrepancies between automated scores and teacher evaluation confirmed that scoring algorithms required professional contextualisation rather than direct instructional application.

These two domains are not independent illustrations but interconnected manifestations of the same principle: analytics generate descriptive information, but its educational significance – whether a pattern signals conceptual misunderstanding, processing overload, or developmental variation – is determined through professional interpretation grounded in classroom evidence. This interpretive work, which analytics cannot substitute, also structured the learning opportunities through which agency was exercised.

Autonomy-supportive engagement as structured and scaffolded agency

Learner agency in this study did not manifest as unrestricted independence, but as structured autonomy within clearly defined instructional boundaries. Students demonstrated agency through pacing decisions, selective review of resources, and strategic engagement with AI-supported feedback. The teacher-designed environment provided scaffolding: video modules enabled selective replay, error logs supported targeted review, and classroom activities offered communicative application. Although students reported positive perceptions, substantial variation existed – learners with more strategic AI feedback use appeared more confident in self-monitoring, while others preferred teacher guidance, indicating that meaningful engagement did not require complete independence.

This interpretation challenges CALL assumptions that associate increasing learner independence with greater autonomy – critiqued by Stockwell and Reinders (2019) and reconceptualised by Godwin-Jones (2024) through distributed agency. In the present study, agency emerged as structured participation within a pedagogically sequenced environment, aligning with sociocultural accounts of autonomy as a socially mediated developmental process (Lantolf & Thorne, 2006; Benson, 2011; Little, 2007).

Student K illustrates both this conceptualisation and its limitations: despite limited participation due to late enrolment, linguistic performance remained within the cohort range, supporting the view that agency can be exercised selectively within accessible domains even when overall participation is constrained; yet analytics may not fully capture learning mechanisms, underscoring the need for complementary qualitative evidence. K's questionnaire responses tell the same story: despite the cohort's lowest platform participation, K rated agency at the scale maximum (M=5.00; see SM1, Section 3.3). Student D illustrates a complementary pattern: despite reserved questionnaire ratings and no open-ended commentary, engagement was the highest in the cohort (weighted replay 467.7%), suggesting that agency may in such cases be expressed through engagement itself rather than through verbal articulation. Pedagogical balance thus requires aligning autonomous practice opportunities with targeted guidance, ensuring that student choice remains goal-directed rather than open-ended.

Human–AI shared regulation as an integrative perspective

The findings suggest that learning regulation cannot be adequately explained through either a fully autonomous or a technology-driven model. Instead, learner development emerged through human–AI shared regulation, in which teachers, learners, and AI systems jointly contributed to regulating learning activities.

AI functions provided immediate feedback, task records, and error information, increasing visibility and self-monitoring opportunities. However, teacher reflection logs consistently showed that AI-generated information alone was insufficient for instructional decision-making; teachers interpreted error

patterns, contextualised feedback, and distinguished between conceptual misunderstanding and cognitive overload. This pattern was particularly evident in pronunciation learning: as documented in Theme 1, automated scores systematically underestimated performance, requiring teacher contextualisation and verification against classroom evidence.

At the learner level, students actively reviewed resources, revisited error logs, and adjusted behaviour in response to both human and technological feedback. Regulation was distributed across human and technological actors rather than located exclusively within either teachers or learners, providing a more adequate account than either technology-centred or learner-centred explanations alone.

Constraints and challenges

Several constraints were observed despite the pedagogical potential of AI-enhanced blended learning.

First, the opacity of automated scoring mechanisms required sustained teacher oversight, particularly when evaluation intersected with communicative competence and spoken language performance. Automated systems occasionally produced assessments that were difficult to interpret pedagogically, making human validation necessary.

Second, AI-supported analytics did not reduce the importance of teacher involvement; they redistributed rather than eliminated professional workload, shifting efforts from routine monitoring towards data interpretation, pedagogical reasoning, and feedback validation.

Third, student variability in digital literacy, engagement orientation, and device access limited the extent to which independent use of AI-supported functions could be assumed.

These observations underscore two broader implications: AI-generated feedback requires teacher validation rather than authoritative instructional application, and structured guidance and equitable support are necessary when integrating AI into language learning environments – rather than assuming that all learners will benefit equally from AI-supported resources.

Limitations and future research

Several limitations should be acknowledged. This study was conducted within a single German as a second foreign language course with a dual-role teacher–researcher; the absence of an independent observer, inter-rater reliability, and member-checking means findings should be interpreted with caution. Positive questionnaire responses may reflect social desirability effects, close teacher–student relationships inherent in small-group embedded case study designs, and ceiling effects; perceptions should be regarded as indicative rather than definitive evidence. The small sample size and context-specificity limit transferability; the study is an exploratory investigation illuminating pedagogical processes rather than establishing generalisable claims.

Future research should extend this work across different languages, proficiency levels, and institutional contexts. Studies with independent observers or externally designed frameworks would test whether observed mediation patterns extend beyond the researcher's approach. Longitudinal designs could examine how agency and shared regulation evolve over time. Methodologically, combining analytics with interviews, stimulated recall, or learner diaries would provide richer understanding of cognitive and interactional dimensions.

Conclusion

As AI-supported tools become increasingly embedded in blended language education, scholarly attention has shifted from technical effectiveness towards pedagogical integration and learner agency within AI-supported environments. Technology was conceptualised as augmentative infrastructure within a teacher-mediated learning ecology – enhancing rather than replacing human instructional capacities. Drawing on platform-generated analytics, item-level assessment records, structured teacher reflection logs, and a post-semester student questionnaire from 11 participants across one semester, the present study examined how instructional visibility, pedagogical mediation, and learner agency interacted within a blended German as a second foreign language course.

The findings reveal an interconnected chain of processes grounded in the data. AI-supported analytics made learning processes visible – exposing recurring grammatical difficulties, low pronunciation accuracy, and selective patterns of engagement intensity despite near-universal task completion. This visibility was interpreted through professional judgement, informing instructional adaptation – redesign of activities for controlled oral practice, sentence decomposition addressing grammatical co-occurrence, and teacher-led modelling following pronunciation discrepancies. Structured opportunities, in turn, supported learner engagement and agency: teacher-created resources were rated more favourably than AI functions, and learners varied between confident self-monitoring and teacher-guidance-oriented strategies. This chain illustrates how AI-enhanced blended learning functions as a mediated instructional system: analytics provide visibility, teachers provide interpretation and adaptation, and learners provide engagement within pedagogically designed structures.

Theoretical contributions

This study contributes to CALL and learning analytics scholarship in three ways.

First, it advances a human–AI shared regulation perspective in which AI supports data collection, pattern detection, feedback generation, and learning visibility, while teachers retain responsibility for diagnosis, interpretation, pedagogical reasoning, and instructional adaptation.

Second, it reconceptualises learner agency in AI-enhanced language learning

as socially and pedagogically mediated – exercised through the interaction of instructional design, teacher guidance, learner choices, and feedback opportunities rather than emerging automatically from access to AI tools.

Third, it demonstrates that the educational value of AI-supported analytics is not inherent in data production or algorithmic feedback, but emerges through teacher interpretation and pedagogical mediation – reinforcing a human-centred understanding of teacher professionalism in data-rich learning environments and highlighting the continuing importance of professional judgement in AI-supported instruction.

Taken together, these contributions advance a human-centred approach to AI-enhanced language education, in which AI expands teachers' access to pedagogically relevant information while professional judgement remains central to educational decision-making and learner engagement is mediated through pedagogically structured participation.

Pedagogical implications

The findings – anchored in observed patterns of high task completion alongside selective replay, teacher-identified discrepancies between automated and manual evaluation, and student-reported strategic use of AI-supported resources – carry several pedagogical implications.

For teachers, the findings highlight the growing importance of interpretive competence in AI-rich environments – as demonstrated when high completion rates required contextualisation against classroom evidence, and when automated pronunciation scores required teacher validation. Teachers must evaluate AI-generated outputs, distinguish behavioural indicators from meaningful learning, and determine when intervention is necessary. For GFL teachers specifically, pronunciation scores should be treated as preliminary indicators rather than definitive assessments; verification should proceed through (a) holistic auditory review of learner recordings alongside classroom observation, (b) contextual evaluation of system-flagged errors against the instructional stage – distinguishing untaught-phoneme errors from genuine phonological difficulty, and (c) prioritising teacher-led modelling for features where ASR unreliability has been documented. AI-supported pronunciation feedback is therefore pedagogically useful when integrated into teacher-guided instruction rather than used as a stand-alone assessment basis. This implication must be qualified, however, by the dual-role context: the observed centrality of teacher interpretation may reflect the instructor-researcher's own pedagogical orientation, and teachers with different instructional philosophies may mediate analytics differently.

At the institutional level, the findings caution against efficiency-driven AI adoption. Sustainable implementation requires investment not only in technological infrastructure but also in pedagogical support, teacher development, and responsible integration strategies – particularly given student variability in digital literacy and device access observed in the present cohort. Within its contextual boundaries, the study has demonstrated that instructional visibility,

pedagogical mediation, and learner agency can function as complementary components of shared regulation – a pattern that, while specific to this GFL context, suggests a broader design principle. As AI capabilities expand, the central challenge will be how institutions and teachers sustain such partnerships while preserving the pedagogical and relational dimensions at the heart of language education – ensuring that professional judgement retains its constitutive role rather than being displaced by automated assessment.

Data availability statement

The supplementary materials (SM1–SM4) provide: SM1, Questionnaire Instrument and Results; SM2, Teacher Reflection Log Excerpts (representative excerpts, selective coding focusing on RQ-relevant segments); SM3, Item-Level Assessment Records; SM4, Qualitative Coding Manual (definitions, examples, decision rules). Platform-exported analytics data are available upon reasonable request but excluded from SMs because: (1) the data are in Chinese, requiring translation; (2) the volume (individual-level traces across 83 instructional sections) is prohibitively large; (3) records contain residual identifiable traces. Key summaries are presented in Tables 3–6 (engagement and performance indicators) and SM3 (item-level assessment records with error-rate summaries).

Authorship contribution statement

The sole author is responsible for all aspects of this work: conceptualisation, methodology, formal analysis, investigation, writing – original draft, writing – review & editing, and project administration.

Funding disclosure statement

This research did not receive any specific funding.

Competing interests statement

The author declares no competing interests.

Ethical statement

This study received retrospective ethics approval from the AI & Cognition Lab Committee on Human Research Protections at Beijing International Studies University (Approval No.: 20260008, dated 6 July 2026). As routine instructional data was not subject to mandatory ethics review at the host institution at the time of collection, approval was sought retrospectively upon journal requirement. Course data derived from routine instructional activities; the questionnaire was voluntary. Retrospective informed consent was obtained, explicitly disclosing that questionnaire responses would be linked to course records through pseudonymised codes and that the instructor-researcher

could potentially identify individual responses. Participation had no influence on grades or academic standing. Analytics and assessment records were analysed solely for research purposes after removing personally identifiable information. Teacher reflection logs documented instructional processes without individual student evaluation beyond normal teaching practice.

Generative AI use disclosure statement

Generative AI tools (ChatGPT and DeepSeek) were used during the preparation of this manuscript solely for limited language polishing and visual formatting support. These tools were used between December 2025 and June 2026. No AI tools were used for data collection, analysis, interpretation, or scholarly decision-making. The author reviewed and edited all content and takes full responsibility for the final manuscript.

References

Alm, A. (2024). Exploring autonomy in the AI wilderness: Learner challenges and choices. *Education Sciences, 14*(12), 1369. https://doi.org/10.3390/educsci14121369

Baum, L., & Günay, G. (Eds.). (2025). *Künstliche Intelligenz in DaF/DaZ.* Frank & Timme.

Benson, P. (2011). *Teaching and researching autonomy* (2nd ed.). Pearson.

Bergdahl, N., Nouri, J., Karunaratne, T., Afzaal, M., & Saqr, M. (2020). Learning analytics for blended learning: A systematic review of theory, methodology, and ethical considerations. *International Journal of Learning Analytics and Artificial Intelligence for Education, 2*(2), 46–79. https://doi.org/10.3991/ijai.v2i2.17887

Biebighäuser, K., & Feick, D. (Eds.). (2020). *Digitale Medien in Deutsch als Fremd- und Zweitsprache.* Erich Schmidt Verlag. https://doi.org/10.37307/b.978-3-503-18292-3

Blin, F. (2004). CALL and the development of learner autonomy: Towards an activity-theoretical perspective. *ReCALL, 16*(2), 377–395. https://doi.org/10.1017/S0958344004000928

Boelens, R., De Wever, B., & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. *Educational Research Review, 22,* 1–18. https://doi.org/10.1016/j.edurev.2017.06.001

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In Proceedings of the Second International Conference on Learning Analytics and Knowledge (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 134–138). ACM. https://doi.org/10.1145/2330601.2330636

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Cukurova, M. (2024). The interplay of learning, analytics and artificial intelligence in education: a vision for hybrid intelligence. *British Journal of Educational Technology, 56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69,* 485–492. https://doi.org/10.1016/j.compedu.2013.04.013

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *The JALT CALL Journal, 16*(2), 107–120. https://doi.org/10.29140/jaltcall.v16n2.273

Dombi, J., Sydorenko, T., & Timpe-Laughlin, V. (2024). Openings and closings in human-human versus human-spoken dialogue system conversations. *Language Learning & Technology, 28*(2), 30–58. https://doi.org/10.64152/10125/73571

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines.* Jossey-Bass.

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology, 25*(1), 4–19. https://doi.org/10.64152/10125/44747

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

Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 333–350). Routledge.

Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL, 25*(2), 165–198. https://doi.org/10.1017/S0958344013000013

Guan, L., Li, S., & Gu, M. M. (2024). AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. *Computers and Education: Artificial Intelligence, 7,* Article 100323. https://doi.org/10.1016/j.caeai.2024.100323

Holec, H. (1981). *Autonomy and foreign language learning.* Pergamon Press. (Original work published 1979, Strasbourg: Council of Europe)

Huang, S., & Cassany, D. (2025). Spanish language learning in the AI era: AI as a scaffolding tool. *Journal of China Computer-Assisted Language Learning.* Advance online publication. https://doi.org/10.1515/jccall-2024-0026

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

Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society, 9*(2), 1–6. https://doi.org/10.1177/20539517221142824

Jehoul, A., Van Nuffel, H., & Schiepers, M. (2025). Blend up: empowering LESLLA learners through blended learning. *ReCALL, 37*(2), 157–173. https://doi.org/10.1017/S0958344024000363

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

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

Lai, C., Zhu, W., & Gong, G. (2015). Understanding the quality of out-of-class English learning. *TESOL Quarterly, 49*(2), 278–308. https://doi.org/10.1002/tesq.171

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development.* Oxford University Press.

Lee, H., & Lee, J. H. (2024). The effects of AI-Guided individualised language learning: A meta-analysis. *Language Learning & Technology, 28*(2), 134–162. https://doi.org/10.64152/10125/73575

Ley, T., Tammets, K., Pishtari, G., Chejara, P., Kasepalu, R., Khalil, M., Saar, M., Iiris, T., Väljataga, T., & Wasson, B. (2023). Towards a partnership of teachers and intelligent learning technology: A systematic literature review of model-based learning analytics. *Journal of Computer Assisted Learning, 39*(5), 1397–1417. https://doi.org/10.1111/jcal.12844

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching, 1*(1), 14–29. https://doi.org/10.2167/illt040.0

Liu, F., Jiang, Y., Lai, C., & Jin, T. (2024). Teacher engagement with automated text simplification for differentiated instruction. *Language Learning & Technology, 28*(2), 163–182. https://doi.org/10.64152/10125/73576

Liu, X., & Xiao, Y. (2024). Chinese university teachers' engagement with generative AI in different stages of foreign language teaching: A qualitative enquiry through the prism of ADDIE. *Education and Information Technologies, 30,* 485–508. https://doi.org/10.1007/s10639-024-13117-9

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology, 50*(6), 2824–2838. https://doi.org/10.1111/bjet.12861

Prinsloo, P. (2017). Fleeing from Frankenstein's monster and meeting Kafka on the way: Algorithmic decision-making in higher education. *E-Learning and Digital Media, 14*(3), 138–163. https://doi.org/10.1177/2042753017731355

Reinders, H., & Benson, P. (2017). Research agenda: Language learning

beyond the classroom. *Language Teaching, 50*(4), 561–578. https://doi.org/10.1017/S0261444817000192

Rösler, D. (2008). Deutsch als Fremdsprache mit digitalen Medien: Versuch einer Zwischenbilanz im Jahr 2008. *Info DaF, 35*(4), 373–389. https://doi.org/10.1515/infodaf-2008-0402

Rösler, D. (2010). Die Funktion von Medien im Deutsch als Fremd- und Deutsch als Zweitsprache-Unterricht. In H.-J. Krumm et al. (Eds.), *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch* (pp. 1199–1214). De Gruyter Mouton.

Rösler, D. (2014). Medialer Wandel, didaktische Konstanz? Zur Entwicklung von Online-Kooperationen am Beispiel der DaF-Studiengänge der Universität Gießen. *Informationen Deutsch als Fremdsprache, 41*(6), 489–501. https://doi.org/10.1515/infodaf-2014-0603

Rösler, D. (2016). CALL und CMC – verschwimmen die Grenzen? In T. Zeyer, S. Stuhlmann, & R. D. Jones (Eds.), *Interaktivität beim Fremdsprachenlernen und -lehren mit digitalen Medien: Hit oder Hype?* (pp. 69–86). Gunter Narr.

Rösler, D. (2018). Lernerautonomie und digitale Medien. In Japanische Gesellschaft für Germanistik (Ed.), *Lernerautonomie und Lernstrategien zwischen Klassenraum und digitaler Welt: Perspektiven auf das Deutschlernen in Japan* (pp. 13–29). Iudicium.

Rösler, D., & Würffel, N. (2010). Blended Learning im Fremdsprachenunterricht. *Fremdsprache Deutsch, 42,* 5–11. https://doi.org/10.37307/j.2194-1823.2010.42.03

Rösler, D., & Würffel, N. (2013). Digitale Medien. In I. Oomen-Welke & B. Ahrenholz (Eds.), *Deutschunterricht in Theorie und Praxis: Deutsch als Fremdsprache* (pp. 252–260). Schneider Hohengehren.

Schwab, G., & Hoffmann, S. (2024). Zur Einführung in den Themenschwerpunkt: Interaktion und Digitalisierung im Fremdsprachenunterricht. *Fremdsprachen Lehren und Lernen (FLuL), 53*(1), 2–10. https://doi.org/10.24053/FLuL-2024-0001

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Stockwell, G., & Reinders, H. (2019). Technology, motivation and autonomy, and teacher psychology in language learning: Exploring the myths and possibilities. *Annual Review of Applied Linguistics, 39,* 40–51. https://doi.org/10.1017/S0267190519000084

Warschauer, M., & Xu, Y. (2024). Generative AI for language learning: Entering a new era. *Language Learning & Technology, 28*(2), 1–4. https://doi.org/10.64152/10125/73569

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand, 14,* Article 100224. https://doi.org/10.1016/j.amper.2025.100224

Wu, X.-Y. (2024). Artificial Intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126,* Article 103498. https://doi.org/10.1016/j.system.2024.103498

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0