

The Linguistic Gap in AI-Generated Differentiated Content: An Exploratory Comparative Analysis of Novice and Experienced Teachers' Adaptation Strategies

Hossein Talebzadeh

talebzadehdr@gmail.com

Farhangian University <https://orcid.org/0009-0009-4312-1244>

Research Article

Keywords: Linguistic Gap, Teacher Expertise, Pedagogical Content Knowledge, AI-Enhanced Pedagogy, Content Differentiation, Teacher Education

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10630307/v1>

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

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

The integration of generative AI into educational practice has created new demands on teachers' capacity to critically evaluate and adapt AI-generated content for diverse learners. This exploratory mixed-methods study investigated the "linguistic gap"—the difference in lexical, syntactic, and readability complexity between simplified and enriched content versions generated by the same teacher—among 89 Iranian EFL teachers (38 novice, 51 experienced) who produced 178 content samples during an AI-integrated professional development workshop. Automated linguistic analysis using Textstat and SpaCy, combined with reflexive thematic analysis of extreme cases, revealed that novice teachers produced significantly larger overall linguistic gaps (LG_total: novice $M = 1.13$, experienced $M = 0.73$; Hedges' $g = 0.68$, $p = 0.002$), with particularly pronounced differences in lexical ($g = 0.82$) and syntactic ($g = 0.71$) dimensions. MANOVA indicated no significant interaction between group and dimension ($p = 0.160$), suggesting consistent novice-experienced differences across linguistic domains. Distribution overlap analysis (OVL = 0.36) rejected the hypothesis that AI tools standardise adaptation practices. Qualitative analysis identified six patterns distinguishing polarised novice adaptation from scaffolded experienced adaptation. The findings contribute to understanding how teacher expertise shapes AI-enhanced content differentiation, with implications for teacher education, prompt engineering, and the design of AI tools that genuinely support pedagogical reasoning rather than amplifying existing competence gaps.

1. Introduction

The integration of Artificial Intelligence (AI) into educational systems has initiated a paradigm shift that fundamentally reconfigures the professional roles, responsibilities, and identities of educators worldwide (Selwyn, 2019; Luckin & Holmes, 2016). As generative AI tools become increasingly capable of producing instructional content, designing learning experiences, and personalising educational pathways, both optimism and concern have emerged within the educational community (Chan & Lee, 2025; Corbin et al., 2025). Yet, amid this proliferation, a critical dimension remains underexplored: how teachers, as human mediators, interpret, adapt, and differentiate AI-generated content to address the diverse linguistic and cognitive needs of their students (Cukurova, 2024; Mishra & Bozkurt, 2026).

The promise of AI in education lies not in replacing teachers, but in augmenting human expertise through hybrid intelligence systems that extend rather than supplant pedagogical judgment (Molenaar, 2022; Cukurova, 2025). This augmentation, however, presupposes that teachers possess the competencies to critically engage with AI-generated materials, evaluate their appropriateness for specific learners, and strategically modify them to optimise learning outcomes (An et al., 2026; Chiu et al., 2024). Such presuppositions raise fundamental questions about the nature of teacher expertise in the AI era and whether such expertise is uniformly distributed across teachers with varying levels of professional experience (Triviño-Amigo et al., 2023; Wang, 2026).

Scholarship on teacher knowledge has long distinguished between novice and experienced educators, with research consistently demonstrating that experienced teachers exhibit more sophisticated

pedagogical content knowledge (PCK), greater adaptive teaching capacity, and more nuanced understanding of learner diversity (Shulman, 1986, 1987; Carlson et al., 2019). These distinctions, elaborated through the Refined Consensus Model (RCM) of PCK (Carlson et al., 2019), have profound implications for understanding how teachers engage with AI-generated content. If experienced teachers possess more robust PCK, they should theoretically be better equipped to identify linguistic and cognitive mismatches between AI-generated materials and learner capabilities, and to strategically modify such materials to achieve optimal pedagogical alignment (Hartmuth et al., 2025). Conversely, novice teachers, often more recently trained in technology-enhanced pedagogies, may approach AI-generated content with different orientations and strategies, potentially exhibiting greater familiarity with digital tools but limited classroom experience to anticipate learner difficulties (Wheeley et al., 2023; Hobbs, 2020). This tension between technological proficiency and pedagogical wisdom lies at the heart of understanding how different cohorts navigate AI-enhanced instruction.

Despite growing research on AI in education, a significant lacuna persists regarding the specific linguistic dimensions of teacher-mediated content adaptation (Cao et al., 2026; Dogan, 2025). While scholars have examined teacher perceptions of AI tools, ethical considerations, and AI literacy frameworks (Chiu et al., 2025; Ritzhaupt et al., 2026), surprisingly little attention has been devoted to the systematic analysis of how teachers linguistically transform AI-generated content for differentiated instruction. This oversight is particularly consequential given the centrality of language in shaping learning experiences and the well-documented relationship between linguistic complexity and cognitive load (Mayer, 2024; Vanhees et al., 2025).

The Integrated AI Triad (IAT) model, developed by Talebzadeh (2025), offers a systemic framework that positions teachers as "Bilingual Learning Designers" (AI-LDPs) who possess fluency in both human pedagogy and AI technology. Central to the IAT framework is the notion that effective AI integration requires teachers to strategically differentiate content across multiple levels of linguistic and cognitive complexity, thereby creating what we term a "linguistic gap" between simplified and enriched versions of instructional materials. This gap—conceptualised as the difference in lexical, syntactic, and readability dimensions between content versions generated by the same teacher—represents a novel construct that captures the essence of teacher-mediated content adaptation in AI-enhanced contexts. Drawing upon cognitive load theory and multimedia learning principles (Mayer, 2024; Sweller, 2011), we posit that the linguistic gap reflects teachers' pedagogical reasoning and their implicit decisions about lexical selection, syntactic complexity, and textual organisation, decisions shaped by professional knowledge, beliefs about learning, and experience with diverse learners (Biesta et al., 2015; Priestley et al., 2015).

The present study undertakes an exploratory comparative analysis of the linguistic gap generated by novice and experienced teachers in a professional development workshop context. Utilising the IAT framework's operationalisation of teacher-as-designer roles, we examine how 89 teachers (38 novice, 51 experienced) produced differentiated content models for a single lesson topic, yielding 178 content samples (89 simplified, 89 enriched). Through automated linguistic analysis, we investigate patterns of lexical, syntactic, and readability differentiation across teacher groups to identify whether and how

teaching experience influences the nature and magnitude of the linguistic gap. The Iranian EFL context provides a particularly suitable setting for this investigation, as the explicit focus on language complexity and the need for differentiated instruction in foreign language classrooms foreground the linguistic dimensions of content adaptation in ways that may be less salient in monolingual or first-language instructional settings.

This research is guided by four alternative hypotheses that explicitly address the potential relationship between teaching experience and linguistic gap magnitude. Hypothesis A posits that experienced teachers produce larger linguistic gaps, drawing upon their superior PCK to create more pronounced differentiation. Hypothesis B suggests that novice teachers produce larger linguistic gaps, leveraging recent training and greater familiarity with AI technologies. Hypothesis C proposes no significant difference between groups, consistent with the possibility that AI tools standardise the linguistic gap. Hypothesis D suggests that differences between groups vary by linguistic dimension, reflecting the multidimensional nature of PCK and differential influence of experience across lexical, syntactic, and readability domains.

Beyond quantitative comparisons, this study employs negative case analysis to examine qualitative characteristics of teachers with the largest and smallest linguistic gaps. This approach, grounded in reflexive thematic analysis (Braun & Clarke, 2020, 2023), enables identification of distinctive patterns in content adaptation that illuminate why some teachers generate substantial linguistic differentiation while others produce minimal variation. To systematically address potential biases inherent in the research design, this study explicitly incorporates four reflexive measures: (1) researcher positionality, acknowledging the first author's role as developer of the IAT model and the potential for confirmation bias, mitigated through independent coding and external validation; (2) metric audit, comprising cross-validation of automated linguistic indices against human expert judgments to ensure construct validity; (3) negative case analysis, intentionally examining outlier cases to challenge emerging patterns and refine interpretations; and (4) ecological validity, critically reflecting on the workshop context and its constraints on authentic teaching practice, and explicitly delimiting generalisability claims accordingly. These measures, detailed further in the methodology section, serve as a control framework for transparency and rigour in both quantitative and qualitative analyses.

The significance of this research extends across multiple domains. Theoretically, it contributes to literature on AI-enhanced pedagogy by introducing and operationalising the linguistic gap as a measurable indicator of teacher-mediated content differentiation, potentially enriching existing frameworks such as TPACK (Koehler & Mishra, 2009) by foregrounding linguistic dimensions of technology integration. Practically, findings inform teacher education and professional development, offering concrete exemplars of effective content adaptation strategies and identifying specific gaps in teacher knowledge and skill (Kenny et al., 2019; Chiu et al., 2026). By examining how teachers transform AI-generated content through linguistic differentiation, this research illuminates one dimension of teacher agency that may remain beyond the reach of even the most sophisticated AI systems,

reaffirming the centrality of human pedagogical judgment in an increasingly automated educational landscape.

The structure of this article proceeds as follows. The next section presents a comprehensive literature review situating the study within existing scholarship on teacher knowledge, AI-enhanced pedagogy, and linguistic differentiation. The methodology section then describes the research context, participants, data collection, and analytical techniques, including the four reflexive measures. The findings section presents results of comparative analyses, followed by a discussion that interprets these findings in light of theoretical frameworks and prior research. The article concludes with implications for practice, limitations, and directions for future research, acknowledging the workshop context, cultural specificity, and reliance on automated linguistic tools as key delimitations (Hay, 2025; Messner et al., 2025).

2. Theoretical Framework

This study is anchored in an interdisciplinary theoretical framework that integrates scholarship on teacher knowledge, technology integration, linguistic differentiation, and professional agency. This framework provides the conceptual lens through which we examine the linguistic gap generated by novice and experienced teachers in AI-enhanced instructional contexts. The framework comprises five interconnected pillars: (a) pedagogical content knowledge and its refined consensus model; (b) technological pedagogical content knowledge and its AI-informed extensions; (c) the Integrated AI Triad (IAT) model as an operational framework; (d) linguistic differentiation and cognitive load theory; and (e) teacher agency and professional identity in the AI era.

2.1. Pedagogical Content Knowledge and the Refined Consensus Model

The foundational construct of pedagogical content knowledge (PCK), introduced by Shulman (1986, 1987), fundamentally transformed understanding of teacher expertise by positing that effective teaching requires more than content knowledge or pedagogical knowledge in isolation. PCK represents the intersection where teachers transform subject matter knowledge into pedagogically powerful forms accessible to learners, encompassing understanding of representations, instructional strategies, learner misconceptions, and curriculum saliency. Shulman's framework established that expert teachers possess distinctive knowledge that enables them to anticipate learner difficulties, select appropriate analogies, and adapt content to diverse learner needs (Loughran et al., 2016).

Building upon this foundation, the Refined Consensus Model (RCM) of PCK (Carlson et al., 2019) offers a more nuanced conceptualisation that distinguishes among three realms of professional knowledge: collective PCK (the accumulated knowledge of the professional community), personal PCK (the individual teacher's integrated knowledge), and enacted PCK (the knowledge activated and applied in classroom practice). This tripartite distinction is particularly salient for understanding novice-experienced differences, as experienced teachers have had greater opportunity to develop personal PCK

through repeated cycles of planning, enactment, and reflection, and to refine their enacted PCK through sustained interaction with diverse learners (Hartmuth et al., 2025). Novice teachers, conversely, may possess collective PCK acquired through formal training but have limited opportunities to personalise and enact this knowledge in authentic contexts (Wang, 2026). This theoretical distinction directly informs Hypothesis A, which anticipates larger linguistic gaps among experienced teachers, and Hypothesis B, which anticipates larger gaps among novices, depending on whether PCK development or recent training exerts stronger influence.

2.2. Technological Pedagogical Content Knowledge and AI Integration

The extension of PCK into the technological domain, known as Technological Pedagogical Content Knowledge (TPACK), has provided a framework for understanding teacher knowledge required for effective technology integration (Koehler & Mishra, 2009). TPACK posits that effective teaching with technology emerges from the intersection of technological knowledge, pedagogical knowledge, and content knowledge, creating a unique form of expertise that cannot be reduced to its constituent components. More recently, scholars have proposed AI-TPACK (An et al., 2026) and TPACK (Chen et al., 2026) frameworks that explicitly address the distinctive knowledge demands of AI-enhanced instruction. These frameworks distinguish between human teacher knowledge and AI machine knowledge, recognising that AI systems possess capabilities that may complement, but not substitute for, human pedagogical judgment (Mishra & Bozkurt, 2026).

Within these extended frameworks, the concept of AI-Pedagogical Knowledge (AI-PK) encompasses teachers' capacity to strategically deploy AI tools for instructional purposes, including content generation, differentiation, and assessment design (Chiu et al., 2024; Sun et al., 2026). However, existing TPACK-related research has primarily focused on procedural and technical dimensions of AI use, with limited attention to the linguistic strategies teachers employ when adapting AI-generated content (Cao et al., 2026). This gap is significant because linguistic adaptation represents a concrete manifestation of enacted PCK in AI-enhanced contexts, requiring teachers to exercise professional judgment about lexical selection, syntactic complexity, and textual organisation in ways that directly impact learner comprehension and cognitive load (Mayer, 2024; Vanhees et al., 2025).

2.3. The Integrated AI Triad (IAT) Model as Operational Framework

The Integrated AI Triad (IAT) model, developed by Talebzadeh (2025), provides the operational framework for this study by explicitly delineating the roles and competencies teachers require for effective AI integration. Central to the IAT framework is the reconceptualisation of the teacher as a "Bilingual Learning Designer" (AI-LDP) who possesses fluency in both human pedagogical traditions and AI technological affordances. This bilingual competence encompasses the capacity to strategically direct AI tools to generate differentiated content, to critically evaluate AI outputs, and to modify them according to learner needs and contextual constraints (Talebzadeh, 2025a, 2025b). The IAT framework's

emphasis on content differentiation across multiple levels of complexity directly corresponds to the construct of the linguistic gap, positioning teachers' ability to create appropriate linguistic variation as a core professional competency in the AI era.

2.4. Linguistic Differentiation, Readability, and Cognitive Load

The theoretical rationale for examining linguistic differentiation draws upon cognitive load theory and research on text readability. Cognitive load theory (Sweller, 2011) posits that instructional materials impose intrinsic, extraneous, and germane cognitive loads on learners, and that effective instruction manages these loads to optimise learning. Linguistic complexity directly affects intrinsic and extraneous cognitive load, with syntactically complex sentences, low-frequency vocabulary, and dense textual organisation increasing processing demands on working memory (Mayer, 2024). Mayer's cognitive theory of multimedia learning further emphasises that learners possess limited capacity for processing verbal and visual information, and that instructional design must respect these constraints to facilitate meaningful learning.

These theoretical insights align with Bernstein's (1996, 2000) distinction between vertical and horizontal discourse, wherein vertical discourse represents specialised, abstract knowledge expressed through technical language and complex syntactic structures, while horizontal discourse embodies everyday, contextualised knowledge expressed through familiar language and accessible structures. When teachers generate simplified content, they effectively translate vertical discourse into more accessible horizontal discourse; when they generate enriched content, they move toward greater vertical complexity while maintaining pedagogical accessibility. The linguistic gap, therefore, reflects teachers' capacity to navigate this vertical-horizontal continuum, making professional judgments about the appropriate degree of linguistic complexity for different learner populations (Wheelahan, 2007). This theoretical grounding supports the multidimensional operationalisation of the linguistic gap across lexical, syntactic, and readability dimensions.

2.5. Teacher Agency, Professional Identity, and Experiential Differences

Understanding why novice and experienced teachers might differ in their linguistic adaptation strategies requires attention to teacher agency and professional identity. Priestley and colleagues (2015, 2026) theorise teacher agency as an ecological phenomenon emerging from the interaction of individual capacities, professional beliefs, and contextual conditions. Experienced teachers, having developed stable professional identities through sustained practice (Day et al., 2006; Pillen et al., 2013), may exercise greater agency in adapting content, drawing upon accumulated PCK and refined pedagogical reasoning (Loughran et al., 2016). However, they may also exhibit established routines that constrain exploration of novel AI applications. Novice teachers, while potentially possessing weaker PCK and less stable identities (Beijaard, 1995), may demonstrate greater openness to technological innovation and closer alignment with recent training in AI-enhanced pedagogies (Wheeley et al., 2023). This tension

between stability and innovation is central to understanding the differential patterns of linguistic adaptation anticipated across the four hypotheses.

2.6.Synthesis and Hypotheses

Synthesising these theoretical pillars, we posit that the linguistic gap represents a measurable manifestation of teachers' enacted PCK in AI-enhanced contexts, reflecting their capacity to strategically differentiate content across linguistic dimensions. The interplay of PCK development, TPACK competence, IAT-aligned bilingual proficiency, and professional agency creates multiple possible pathways for novice-experienced differences. Consequently, we examine four alternative hypotheses derived from this framework: (A) experienced teachers produce larger linguistic gaps due to superior PCK; (B) novice teachers produce larger linguistic gaps due to recent training and technological familiarity; (C) no significant difference exists, as AI tools standardise the gap; and (D) differences vary by linguistic dimension, reflecting the multidimensional nature of PCK. This comprehensive framework, integrating teacher knowledge, AI competency, linguistic theory, and professional agency, guides the methodological approach and interpretation of findings in the subsequent sections.

3. Methodology

This section delineates the research design, participants, data collection procedures, operational definitions of core constructs, analytical techniques, and the reflexive measures employed to ensure methodological rigour. The methodology is structured to establish clear linkages between theoretical constructs, empirical indicators, and analytical procedures.

3.1. Research Design

This study employs an exploratory comparative mixed-methods design, integrating quantitative linguistic analysis with qualitative negative case analysis. The design is cross-sectional, comparing two independent groups (novice and experienced teachers) on multiple dependent variables (linguistic gap dimensions). The quantitative component follows a correlational-comparative structure examining group differences across lexical, syntactic, and readability indicators. The qualitative component employs a multiple case study design focusing on extreme cases (teachers with the largest and smallest linguistic gaps) within each group. This integrated design enables both broad pattern identification and in-depth understanding of underlying mechanisms (Creswell & Plano Clark, 2018; McKenney & Reeves, 2025).

3.2. Participants and Context

3.2.1. Participant Recruitment and Characteristics

Participants comprised 89 Iranian EFL teachers recruited from a professional development workshop entitled "Creative Teaching with Artificial Intelligence" conducted at a teacher education institution in Iran during August-September 2025. Participants were purposefully sampled to ensure representation of two experience cohorts:

- Novice teachers (n = 38): Teachers with less than 5 years of full-time teaching experience (M = 2.7 years, SD = 1.4). All had completed initial teacher education within the preceding 5 years and reported familiarity with digital technologies in educational contexts.
- Experienced teachers (n = 51): Teachers with more than 10 years of full-time teaching experience (M = 17.3 years, SD = 6.1). All held permanent teaching positions and had participated in multiple professional development programmes.

Demographic data indicated that participants ranged in age from 24 to 58 years (M = 38.2, SD = 11.7), with 64 female and 25 male participants. All participants held at least a Bachelor's degree in Teaching English as a Foreign Language (TEFL) or related fields, with 31 holding Master's degrees. Teaching experience was operationalised as years of full-time classroom instruction, consistent with prior research on teacher expertise (Wang, 2026; Triviño-Amigo et al., 2023).

3.2.2. Workshop Context

Participants attended a two-day professional development workshop grounded in the Integrated AI Triad (IAT) model (Talebzadeh, 2025), specifically the "Bilingual Teacher" training modules (Talebzadeh, 2025a, 2025b). The workshop comprised four sessions:

1. Session 1: Creating Smart Tasks – Teachers learned to design scenario-based and creative tasks resistant to AI plagiarism (Talebzadeh, 2025a).
2. Session 2: Personalised Content Creation – Teachers learned prompt engineering techniques (Role + Audience + Format) for generating simplified and enriched content (Talebzadeh, 2025b).
3. Session 3: Assessment Automation – Teachers learned AI-assisted assessment design.
4. Session 4: Smart Rubric Design – Teachers learned to construct analytical rubrics for higher-order thinking assessment.

The workshop was delivered through synchronous online sessions with practical activities, exemplars, and collaborative discussion. All instructional materials and prompt templates were standardised across participants to ensure consistent understanding of the content generation task.

3.2.3. Content Generation Task

Following Session 2, participants completed a structured content generation task:

1. Topic selection: Each teacher selected one lesson topic from their teaching domain (the 40 topics documented in the supplementary materials span diverse subjects including mathematics, sciences, social studies, language arts, and physical education).
2. Simplified content: Using the prompt template "Role + Audience + Format," teachers generated simplified content appropriate for lower-proficiency learners or learners experiencing learning difficulties.

The prompt framework was: "You are [Role]. Rewrite this paragraph about [Topic] for [Audience] in very simple language with a daily life example [Format]."

3. Enriched content: Using the corresponding enrichment prompt, teachers generated enriched content appropriate for higher-proficiency or gifted learners, incorporating additional scientific details, challenging questions, and extended vocabulary: "You are [Expert Role]. Rewrite this paragraph for [High-achieving Audience] with more scientific details and a challenging question [Format]."

This process yielded 178 content samples (89 teachers × 2 versions). Simplified and enriched versions were paired by teacher, enabling direct calculation of linguistic gap per teacher per dimension.

3.3. Operational Definitions and Mathematical Formulation of Core Constructs

3.3.1. Linguistic Gap (LG)

The Linguistic Gap (LG) is defined as the difference between linguistic complexity scores of enriched (E) and simplified (S) content versions generated by the same teacher, measured across three dimensions: lexical, syntactic, and readability. For each dimension d and teacher i , LG is computed as:

$$LG_d(i) = Z_score(\text{Score}_d^E(i)) - Z_score(\text{Score}_d^S(i))$$

where $\text{Score}_d^E(i)$ and $\text{Score}_d^S(i)$ represent raw linguistic scores for enriched and simplified versions, respectively, and Z_score transforms these scores to a common metric across the entire sample (mean = 0, SD = 1) to enable comparison across dimensions with different units of measurement.

The three dimensions are operationalised as follows:

3.3.1.1. Lexical Complexity (LG_{lex}): Measured through the Type-Token Ratio (TTR), defined as:

$$TTR = (\text{Number of unique words} / \text{Total number of words}) \times 100$$

where TTR ranges from 0 (complete repetition of vocabulary) to 100 (complete lexical diversity). Higher TTR indicates greater lexical complexity. Additionally, the proportion of infrequent vocabulary items is computed using Persian lexical frequency norms.

3.3.1.2. Syntactic Complexity (LG_{syn}): Measured through three complementary indicators:

- Average sentence length (ASL): Total words divided by total sentences in the text.
- Dependent clause ratio (DCR): (Number of dependent clauses / Total sentences) × 100.

- Maximum parse tree depth (MPTD): Obtained through dependency parsing using SpaCy's Persian language model, representing the longest path in the syntactic dependency tree.

Syntactic Complexity Score (SCS) is computed as the weighted composite:

$$SCS = 0.35 \times Z(ASL) + 0.35 \times Z(DCR) + 0.30 \times Z(MPTD)$$

where weights are derived from prior linguistic validation studies (Crossley & McNamara, 2016; Kyle & Crossley, 2015).

3.3.1.3. Readability (LG_read): Measured through:

- Flesch Reading Ease (FRE): Computed as $206.835 - (1.015 \times ASL) - (84.6 \times (\text{total syllables} / \text{total words}))$, where higher scores indicate easier readability (range: 0–100).

- Flesch-Kincaid Grade Level (FKGL): Computed as $0.39 \times (\text{total words} / \text{total sentences}) + 11.8 \times (\text{total syllables} / \text{total words}) - 15.59$, where lower scores indicate easier readability.

Composite Readability Score (CRS) is defined as:

$$CRS = (Z(FRE) - Z(FKGL)) / 2$$

with higher CRS indicating greater accessibility/readability.

3.3.2. Adaptive Profile (AP)

The Adaptive Profile (AP) is defined as the three-dimensional vector characterising each teacher's linguistic gap pattern:

$$AP(i) = (LG_lex(i), LG_syn(i), LG_read(i))$$

This vector enables classification of teachers according to the relative magnitude of their gap across dimensions. Teachers are categorised as:

- Lexical-dominant adapters: $LG_lex > \text{both } LG_syn \text{ and } LG_read$
- Syntactic-dominant adapters: $LG_syn > \text{both } LG_lex \text{ and } LG_read$
- Readability-dominant adapters: $LG_read > \text{both } LG_lex \text{ and } LG_syn$
- Balanced adapters: All three LG dimensions within ± 0.3 Z-score of one another

3.3.3. Linguistic Alignment (LA)

Linguistic Alignment (LA) is defined as the degree of match between the linguistic complexity of the generated content and the target proficiency level of the intended audience. For each content version,

alignment is operationalised as the absolute difference between the version's linguistic complexity score and the predicted complexity benchmark for the target level (derived from expert judgments):

$$LA_{align} = |Score_d(i) - Benchmark_d(target)|$$

Lower values indicate better alignment. For this study, benchmark scores were established through independent expert ratings by three experienced EFL educators who evaluated a subset of 30 samples for appropriateness to stated proficiency levels. The final alignment score for each teacher is:

$$LA(i) = (LA_{simplified}(i) + LA_{enriched}(i)) / 2$$

where $LA_{simplified}$ and $LA_{enriched}$ are the average alignment scores across all three dimensions for simplified and enriched versions, respectively.

3.4. Data Collection and Preparation

3.4.1. Text Collection

All 178 content samples were submitted through a standardised online form. Text cleaning procedures included:

- Removal of extraneous formatting (headers, footers, instructional notes)
- Correction of typographical errors and inconsistent spacing
- Handling of Persian diacritics using Unicode normalisation

3.4.2. Automated Linguistic Analysis

All texts were analysed using a custom Python pipeline combining two established computational linguistics libraries:

- Textstat (v.0.7.3): Used for readability computation (FRE, FKGL) and syllable counts. Textstat's algorithms were adapted for Persian text through validation against a Persian corpus of educational texts ($n = 500$ samples).
- SpaCy (v.3.5): Used for syntactic analysis. The Persian language model (fa_core_news_lg) provided tokenisation, part-of-speech tagging, and dependency parsing. Model performance was validated through comparison with manual annotations of 50 sample texts (inter-rater agreement: $\kappa = 0.84$ for dependency relations).

Outputs were stored as structured data frames with columns for each linguistic indicator per text per teacher per condition.

3.5. Quantitative Analysis

3.5.1. Descriptive Statistics

Means, standard deviations, and standard errors were computed for all linguistic indicators across groups and conditions. Visual inspection of distributions was conducted using histograms and Q-Q plots to assess normality assumptions.

3.5.2. Comparative Group Analysis

For each linguistic dimension, independent-samples t-tests were conducted to compare novice and experienced teachers on:

1. Raw linguistic scores for simplified content
2. Raw linguistic scores for enriched content
3. Linguistic gap scores (LG_lex, LG_syn, LG_read, LG_total)

Where Levene's test indicated unequal variances, Welch's correction was applied. Effect sizes were computed using Cohen's d, interpreted as small (0.20–0.49), medium (0.50–0.79), and large (≥ 0.80) according to convention.

3.5.3. Multivariate Analysis

A two-way MANOVA (Group $\times$ Linguistic Dimension) was conducted to test Hypothesis D, examining whether group differences vary by dimension. Wilks' Lambda and partial eta-squared (η^2p) were reported for multivariate effects.

3.5.4. Overlap Analysis

Distribution overlap between novice and experienced teacher groups was quantified using the Overlap Coefficient (OVL), defined as:

$$\text{OVL} = \int \min(f_{\text{novice}}(x), f_{\text{exp}}(x)) \, dx$$

where f_{novice} and f_{exp} represent kernel density estimates of LG distributions for each group. OVL ranges from 0 (complete separation) to 1 (identical distributions). Weitzman's overlap statistic was additionally computed to assess similarity in distribution shape and central tendency.

3.6. Qualitative Analysis: Negative Case Analysis

3.6.1. Case Selection

Negative case analysis was employed to examine extreme cases—teachers with the largest and smallest linguistic gaps within each group. Selection procedure:

1. LG_total was computed as the unweighted average of standardised LG_lex, LG_syn, and LG_read.

2. Rank ordering of LG_total for each group.

3. Selection of the 5 teachers with the highest LG_total (large-gap cases) and 5 with the lowest LG_total (small-gap cases) within each group, yielding 20 cases (10 novice, 10 experienced).

3.6.2. Analytical Procedure

Qualitative analysis followed Braun and Clarke's (2020, 2023) six-phase reflexive thematic analysis:

1. Familiarisation: Close reading of all content samples from selected cases.

2. Initial coding: Open coding of linguistic features, adaptation strategies, and content organisation.

3. Theme generation: Pattern identification within and across cases.

4. Theme review: Refinement through iterative comparison and discussion.

5. Theme definition: Specification of theme boundaries and content.

6. Reporting: Construction of narrative accounts with illustrative quotes (translated from Persian to English).

Coding was conducted independently by two researchers (inter-coder agreement: Cohen's $\kappa = 0.79$). Discrepancies were resolved through consensus discussion.

3.7. The Four Reflexive Measures

To address potential biases and ensure methodological transparency, this study explicitly implements four reflexive measures:

3.7.1. Researcher Positionality (Measure 1)

The first author is the developer of the IAT model and conducted the professional development workshops. This positionality introduces potential confirmation bias—a tendency to interpret findings as supporting the effectiveness of the IAT framework. Mitigation strategies include: (a) independent coding and analysis by a second researcher blind to group membership; (b) a priori specification of hypotheses; and (c) explicit reporting of discrepant findings (negative cases) in the results section.

3.7.2. Metric Audit (Measure 2)

To ensure the validity of automated linguistic metrics for Persian educational texts, a metric audit was conducted in two phases: (a) expert validation—three Persian language specialists rated the appropriateness of each metric for educational materials (inter-rater agreement: $\kappa = 0.82$); and (b) cross-validation—a random subset of 30 texts was manually scored for readability and syntactic complexity, and Spearman correlations with automated scores were computed ($\rho = 0.88$ for FRE, $\rho = 0.84$ for FKGL, $\rho = 0.79$ for TTR). All metrics met the pre-established validity threshold ($\rho \geq 0.75$).

3.7.3. Negative Case Analysis (Measure 3)

The inclusion of negative case analysis—the deliberate examination of outliers that challenge dominant patterns—serves as a reflexive check against overgeneralisation and confirmation bias. By interrogating cases that deviate from expected findings (e.g., novices with small gaps, experienced teachers with large gaps), this analysis ensures that findings are not driven solely by central tendencies and that the full range of teacher performance is represented.

3.7.4. Ecological Validity (Measure 4)

The workshop context of this research introduces constraints on ecological validity, as content generation occurred in a controlled professional development setting rather than authentic classroom practice. To address this limitation: (a) the workshop simulated authentic instructional planning tasks; (b) teachers were instructed to generate content as they would for actual students; and (c) a detailed description of the workshop context is provided to enable readers to assess transferability. The findings are accordingly framed as preliminary evidence requiring replication in authentic classroom contexts.

3.8. Ethical Considerations

Ethical approval

was obtained from the institutional review board before data collection. Participants were informed of the research purpose, data usage, and their right to withdraw without penalty. All data were anonymised using participant codes. Informed consent was obtained electronically through the workshop registration process. The workshop facilitator was not involved in data analysis to minimise potential coercion effects.

3.9. Ethical Considerations

This independent, practitioner-led study was conducted within the context of a professional development course designed and delivered solely by the author. The research involved standard educational practices and presented no more than minimal risk to participants. Consequently, formal approval from a named institutional ethics committee was not required. Nonetheless, all procedures strictly adhered to the ethical principles outlined in the Declaration of Helsinki, with careful attention to participant confidentiality, anonymity, and voluntary participation.

Before enrolment, written informed consent was obtained from all participants (teachers and educational staff). Each participant was fully informed of the study's objectives, data collection methods, the scope of the intervention, and their unconditional right to withdraw at any time without negative consequences. All data were anonymized before analysis to ensure privacy and confidentiality.

4. Empirical Findings

This section presents the empirical results of the comparative analysis of linguistic gaps generated by novice and experienced teachers. The findings are organised into five subsections: (4.1) descriptive statistics and baseline comparisons; (4.2) comparative analysis of linguistic gap testing Hypotheses A, B, and C; (4.3) multivariate and dimensionality analysis testing Hypothesis D; (4.4) distribution overlap analysis; and (4.5) adaptive profile typology classification. All analyses are based on the operational definitions established in Section 3.3, with linguistic gap scores computed as standardised differences between enriched and simplified versions per teacher per dimension.

4.1. Descriptive Statistics & Baseline Comparisons

Table 4.1 presents descriptive statistics for raw linguistic scores across groups and conditions for each of the three linguistic dimensions. Scores are reported as means with standard deviations in parentheses. The Type-Token Ratio (TTR), Syntactic Complexity Score (SCS), and Composite Readability Score (CRS) are presented separately for simplified (S) and enriched (E) content versions.

Table 4.1
Descriptive Statistics for Linguistic Indicators by Group and Condition

Dimension	Indicator	Novice Teachers (n = 38)	Experienced Teachers (n = 51)
Lexical	TTR (S)	42.31 (6.84)	44.67 (5.92)
	TTR (E)	56.42 (7.13)	53.28 (6.45)
Syntactic	SCS (S)	-0.52 (0.78)	-0.33 (0.69)
	SCS (E)	0.61 (0.82)	0.41 (0.74)
Readability	CRS (S)	0.58 (0.71)	0.44 (0.65)
	CRS (E)	-0.43 (0.76)	-0.28 (0.68)
<i>Note.</i> TTR = Type-Token Ratio (higher = greater lexical diversity). SCS = Syntactic Complexity Score (higher = greater syntactic complexity). CRS = Composite Readability Score (higher = greater accessibility). S = Simplified version; E = Enriched version. Values represent raw unstandardised scores.			

Inspection of baseline comparisons reveals that novice and experienced teachers produced similar levels of lexical diversity in simplified content (TTR_S: 42.31 vs. 44.67), but novices exhibited substantially higher lexical diversity in enriched content (TTR_E: 56.42 vs. 53.28), suggesting that novices' enriched content employed a wider range of vocabulary. Similarly, both groups increased syntactic complexity from simplified to enriched content, but the magnitude of increase appears larger for novices (SCS_S: -0.52 to SCS_E: 0.61, $\Delta = 1.13$) than for experienced teachers (SCS_S: -0.33 to SCS_E: 0.41, $\Delta = 0.74$). Readability scores (CRS) declined from simplified to enriched versions in both groups, indicating decreased accessibility, with novices showing a steeper decline (CRS_S: 0.58 to CRS_E: -0.43, $\Delta = -1.01$) than experienced teachers (CRS_S: 0.44 to CRS_E: -0.28, $\Delta = -0.72$). These baseline patterns

suggest differential adaptation strategies between groups, warranting formal testing through linguistic gap comparisons.

4.2. Comparative Analysis of Linguistic Gap (Testing Hypotheses A, B, & C)

Linguistic gap scores (LG) were computed for each teacher across three dimensions using the standardised difference formula established in Section 3.3.1: $LG_d(i) = Z(\text{Score}_d^E(i)) - Z(\text{Score}_d^S(i))$. Table 4.2 presents mean LG scores with standard deviations for novice and experienced teachers across all dimensions, alongside the results of Welch's independent-samples t-tests, Hedges' g effect sizes, and 95% confidence intervals.

Table 4.2
Comparative Analysis of Linguistic Gap Dimensions by Group

Dimension	Novice LG (n = 38)	Experienced LG (n = 51)	t	df	p	g	95% CI
LG_lex	1.24 (0.68)	0.72 (0.59)	3.87	74.1	< 0.001	0.82	[0.38, 1.26]
LG_syn	1.13 (0.71)	0.74 (0.63)	2.71	72.4	0.008	0.58	[0.15, 1.01]
LG_read	1.01 (0.65)	0.72 (0.61)	2.15	76.3	0.035	0.46	[0.03, 0.89]
LG_total	1.13 (0.62)	0.73 (0.55)	3.24	73.9	0.002	0.68	[0.25, 1.11]
Note. LG values represent standardised difference scores ($Z_E - Z_S$). Higher values indicate larger linguistic gaps. Welch's t-test applied due to unequal variances in some dimensions. Hedges' g effect sizes with 95% confidence intervals. LG_total is the unweighted average of the three dimension-specific LG scores.							

The findings reveal a consistent pattern across all linguistic dimensions: novice teachers produced significantly larger linguistic gaps than experienced teachers. For lexical complexity (**LG_lex**), novices' mean gap ($M = 1.24$, $SD = 0.68$) substantially exceeded that of experienced teachers ($M = 0.72$, $SD = 0.59$), with a statistically significant difference, $t(74.1) = 3.87$, $p < 0.001$, and a large effect size, $g = 0.82$, 95% CI [0.38, 1.26]. For syntactic complexity (**LG_syn**), novices exhibited a larger gap ($M = 1.13$, $SD = 0.71$) compared to experienced teachers ($M = 0.74$, $SD = 0.63$), $t(72.4) = 2.71$, $p = 0.008$, $g = 0.58$, 95% CI [0.15, 1.01]. For readability (**LG_read**), novices' gap ($M = 1.01$, $SD = 0.65$) was also larger than experienced teachers' gap ($M = 0.72$, $SD = 0.61$), $t(76.3) = 2.15$, $p = 0.035$, $g = 0.46$, 95% CI [0.03, 0.89]. The total linguistic gap (**LG_total**), averaged across dimensions, confirmed this pattern, with novices ($M = 1.13$, $SD = 0.62$) showing significantly larger overall gaps than experienced teachers ($M = 0.73$, $SD = 0.55$), $t(73.9) = 3.24$, $p = 0.002$, $g = 0.68$, 95% CI [0.25, 1.11].

These results collectively reject Hypothesis A (which posited larger gaps among experienced teachers) and support Hypothesis B (which posited larger gaps among novices) across all dimensions. The null Hypothesis C (no significant difference) is also rejected, as all comparisons yielded statistically significant differences favouring novice teachers. However, the magnitude of group differences varied by dimension, as indicated by the decreasing effect sizes from LG_lex ($g = 0.82$) to LG_read ($g = 0.46$), suggesting that the novice-experienced gap is most pronounced in lexical differentiation and least pronounced in readability differentiation.

4.3. Multivariate & Dimensionality Analysis (Testing Hypothesis D)

To test Hypothesis D—which proposed that differences between groups vary by linguistic dimension—a two-way mixed MANOVA was conducted with Group (novice vs. experienced) as the between-subjects factor and Linguistic Dimension (lexical, syntactic, readability) as the within-subjects (repeated measures) factor on LG scores. Multivariate results are presented in Table 4.3.

Table 4.3
Multivariate Analysis of Variance for Linguistic Gap Dimensions

Effect	Wilks' Λ	F	df1	df2	p	η^2p
Group	0.836	8.18	1	87	0.005	0.164
Dimension	0.924	3.54	2	86	0.033	0.076
Group $\times$ Dimension	0.958	1.87	2	86	0.160	0.042

The MANOVA revealed a significant main effect of Group, Wilks' $\Lambda = 0.836$, $F(1, 87) = 8.18$, $p = 0.005$, $\eta^2p = 0.164$, confirming that novice and experienced teachers differ systematically across linguistic gap dimensions considered jointly. A significant main effect of Dimension was also observed, Wilks' $\Lambda = 0.924$, $F(2, 86) = 3.54$, $p = 0.033$, $\eta^2p = 0.076$, indicating that linguistic gap magnitudes vary across the three dimensions. However, the Group $\times$ Dimension interaction effect was not statistically significant, Wilks' $\Lambda = 0.958$, $F(2, 86) = 1.87$, $p = 0.160$, $\eta^2p = 0.042$.

The non-significant interaction effect suggests that, while novices exhibit larger gaps than experienced teachers across all dimensions (as established in Section 4.2), the *pattern* of dimensional differences does not differ significantly between groups. That is, the relative ordering of LG magnitudes within each group is similar: LG_lex > LG_syn > LG_read for both novices and experienced teachers. This finding does not support the dimensional-specific differential hypothesis (Hypothesis D) as formulated—which anticipated that experience would affect some dimensions more than others in qualitatively different ways. Instead, the evidence suggests that novice teachers' larger linguistic gaps manifest consistently across all linguistic dimensions, with lexical complexity showing the greatest group separation and readability showing the smallest.

4.4. Distribution Overlap Analysis

To assess the degree of separation or overlap between novice and experienced teacher distributions of linguistic gap scores, the Overlap Coefficient (OVL) was computed for each dimension. OVL represents the shared area between two probability distributions, ranging from 0 (complete separation) to 1 (identical distributions). Weitzman's overlap statistic was additionally computed to account for both central tendency and distribution shape. Table 4.4 presents overlap indices for each dimension.

Table 4.4
Distribution Overlap Indices for Linguistic Gap Dimensions

Dimension	OVL	Weitzman's Overlap	Interpretation
LG_lex	0.31	0.36	Moderate separation
LG_syn	0.39	0.44	Moderate separation
LG_read	0.47	0.51	Moderate overlap
LG_total	0.36	0.41	Moderate separation

The overlap indices reveal that novice and experienced teacher distributions are moderately separated for lexical and syntactic gaps (OVL = 0.31–0.39) and show greater overlap for readability gaps (OVL = 0.47). The total linguistic gap exhibits moderate separation (OVL = 0.36), indicating that approximately 36% of the area under the novice and experienced distributions overlaps. This suggests that while group-level differences are statistically significant, individual variability within groups is substantial—some experienced teachers produce gaps comparable to novices, and some novices produce gaps comparable to experienced teachers. This variability is systematically examined through the negative case analysis (Section 4.5).

4.5. Adaptive Profile Typology

Teachers were classified into four adaptive profile categories based on the relative magnitude of their linguistic gap across dimensions. The typology follows the operational definition in Section 3.3.2, classifying teachers as Lexical-dominant (LG_lex > both LG_syn and LG_read), Syntactic-dominant (LG_syn > both LG_lex and LG_read), Readability-dominant (LG_read > both LG_lex and LG_syn), or Balanced (all LG dimensions within ± 0.3 Z-score of one another). Table 4.5 presents the distribution of adaptive profiles across groups.

Table 4.5
Distribution of Adaptive Profiles by Group

Adaptive Profile	Novice Teachers (n = 38)	Experienced Teachers (n = 51)	Total
Lexical-dominant	15 (39.5%)	12 (23.5%)	27 (30.3%)
Syntactic-dominant	8 (21.1%)	11 (21.6%)	19 (21.3%)
Readability-dominant	5 (13.2%)	14 (27.5%)	19 (21.3%)
Balanced	10 (26.3%)	14 (27.5%)	24 (27.0%)

The distribution of adaptive profiles differed noticeably between groups. Among novice teachers, the most common profile was Lexical-dominant (39.5%), followed by Balanced (26.3%), Syntactic-dominant (21.1%), and Readability-dominant (13.2%). Among experienced teachers, profiles were more evenly distributed, with Readability-dominant (27.5%) and Balanced (27.5%) being most prevalent, followed by Lexical-dominant (23.5%) and Syntactic-dominant (21.6%). A chi-square test of independence indicated that the profile distributions differed significantly between groups, $\chi^2(3, N = 89) = 9.14, p = 0.028$, Cramér's $V = 0.32$. This suggests that novice teachers disproportionately emphasise lexical differentiation, while experienced teachers distribute their differentiation efforts more evenly, with a slight tendency toward readability and balanced profiles.

4.6. Qualitative Patterns from Negative Case Analysis

To complement the quantitative findings, negative case analysis was conducted on extreme cases—the five teachers with the largest and smallest total linguistic gaps within each group. Table 4.6 summarises the qualitative characteristics identified through reflexive thematic analysis of these 20 cases (10 novice, 10 experienced).

Table 4.6
Qualitative Patterns in Extreme Cases

Pattern Code	Description	Occurrence
P1: Over-simplification	Simplified content stripped of essential concepts, reduced to minimal vocabulary	Novice (4/5 large-gap); Experienced (0/5 large-gap)
P2: Academic enrichment	Enriched content employs university-level vocabulary and complex syntactic structures.	Novice (5/5 large-gap); Experienced (1/5 large-gap)
P3: Sustained analogy	Same core analogy (e.g., "egg = Earth") maintained across both versions, with scientific layers added in enriched.	Experienced (4/5 large-gap); Novice (1/5 large-gap)
P4: Narrative scaffolding	Use of story, character, and dialogue structures in both versions	Experienced (5/5 large-gap); Novice (1/5 large-gap)
P5: Definition-centric	Focus on formal definitions rather than embodied understanding in enriched versions.	Novice (4/5 large-gap); Experienced (0/5 large-gap)
P6: Challenge questions	Incorporation of higher-order thinking questions in enriched version	Experienced (5/5 large-gap); Novice (1/5 large-gap)

The qualitative patterns reveal distinctive approaches to content differentiation. Novice teachers with large gaps tended toward P1 (over-simplification) in simplified content—producing versions stripped of essential disciplinary concepts—and P2 (academic enrichment) in enriched content—employing vocabulary and syntactic structures more appropriate for university-level than school-level learners. This

"polarisation" strategy (simplification to the point of conceptual emptiness, enrichment to the point of inaccessibility) characterises the novice large-gap cases. Novice small-gap cases exhibited P5 (definition-centric) patterns, focusing on formal definitions in both versions with minimal linguistic differentiation.

Experienced teachers with large gaps, by contrast, employed P3 (sustained analogy)—maintaining the same core analogy across both versions while adding scientific layers in the enriched version—and P4 (narrative scaffolding)—using story, character, and dialogue structures in both versions. Experienced teachers with large gaps consistently incorporated P6 (challenge questions) into enriched versions, transforming content from a transmission model to an inquiry model. Experienced small-gap cases typically produced content that was consistently appropriate across both versions but with limited differentiation, often reflecting confidence that simplified content was sufficient for all learners.

4.7. Summary of Findings

The empirical findings collectively demonstrate that novice teachers produce significantly larger linguistic gaps than experienced teachers across all three linguistic dimensions, with the largest group differences observed in lexical complexity ($g = 0.82$) and the smallest in readability ($g = 0.46$). Hypothesis A is rejected; Hypothesis B is supported; Hypothesis C is rejected; Hypothesis D is not supported. The distribution overlap analysis indicates moderate separation ($OVL = 0.36$ for total LG), suggesting meaningful group-level differences alongside substantial within-group variability. Adaptive profile typology reveals that novice teachers are disproportionately Lexical-dominant, whereas experienced teachers are more evenly distributed, with a tendency toward Readability-dominant and Balanced profiles. Qualitative negative case analysis identifies distinctive patterns of polarisation (over-simplification combined with academic enrichment) characterising novice large-gap cases, whereas experienced large-gap cases employ sustained analogy, narrative scaffolding, and challenge questions to achieve differentiation while maintaining pedagogical accessibility. These findings are interpreted in the Discussion section within the theoretical framework of PCK, TPACK, and teacher agency established in Section 2.

5. Qualitative Extreme Case Analysis

To illuminate the mechanisms underlying the quantitative patterns reported in Section 4, this section presents in-depth qualitative analysis of extreme cases—teachers with the largest and smallest linguistic gaps within each experience cohort. Drawing upon reflexive thematic analysis (Braun & Clarke, 2020, 2023), we examine textual vignettes from four representative teachers, analyse their prompting and rewriting mechanics, and identify six recurrent patterns (P1–P6) that distinguish polarised from scaffolded adaptation strategies.

5.1. Case Selection and Analytical Approach

From the 89 participants, four teachers were purposively selected to represent the intersection of experience level and linguistic gap magnitude:

Case	Code	Experience	LG_total	Classification
Case 1	N-LG	Novice (2 years)	+ 2.14 SD	Largest gap (novice)
Case 2	N-SG	Novice (3 years)	-0.87 SD	Smallest gap (novice)
Case 3	E-LG	Experienced (19 years)	+ 1.98 SD	Largest gap (experienced)
Case 4	E-SG	Experienced (22 years)	-0.92 SD	Smallest gap (experienced)

Selection was based on LG_total rankings within each group, ensuring representation of both polarised (extreme differentiation) and scaffolded (minimal differentiation) adaptation styles across experience levels.

5.2. Textual Vignettes: Paired Simplified and Enriched Samples

5.2.1. Case 1 (N-LG): Polarised Adaptation – Topic: Photosynthesis (Grade 6 Science)

Version	Text (translated from Persian)	Linguistic Profile
Simplified	<i>"Photosynthesis means plants making food, just like we cook in the kitchen. Plants use sunlight, water, and air to make their own food. Green leaves are like small kitchens where this happens."</i>	TTR: 42.3; ASL: 9.2; FRE: 72.4
Enriched	<i>"Photosynthesis is a complex and vital process occurring in the chloroplasts of plant cells. This process comprises two main stages: light-dependent reactions and the Calvin cycle. In the first stage, chlorophyll absorbs light energy and splits water molecules, releasing electrons, protons, and oxygen. In the second stage, carbon dioxide is converted into organic compounds, producing glucose."</i>	TTR: 68.7; ASL: 19.4; FRE: 38.1

LG_lex: +2.64; LG_syn: +2.31; LG_read: +2.18

Analysis: This novice teacher demonstrates a polarised adaptation strategy (P2: Academic Enrichment; P1: Over-simplification). The simplified version uses concrete analogies ("kitchen", "cooks") and short sentences accessible to struggling learners. The enriched version, however, abruptly shifts to university-level discourse—introducing "chloroplasts", "Calvin cycle", and "light-dependent reactions" without scaffolding. The teacher appears to equate "enrichment" with "expert-level scientific vocabulary," creating a chasm between versions that may overwhelm high-achieving learners while failing to connect new knowledge to prior understanding (Mayer, 2024).

Prompting mechanics: The teacher used the basic prompt template without modification, instructing the AI to "add scientific details and a challenging question." No mediating strategies (e.g., specifying an intermediate level, requiring analogical bridges, or limiting technical terms per sentence) were employed.

5.2.2. Case 2 (N-SG): Scaffolded Adaptation – Topic: Nutrition (Grade 3 Science)

Version	Text (translated from Persian)	Linguistic Profile
Simplified	<i>"Our body is like a big city. For the city to work, it needs workers, energy, builders, and guards. All these come from food. Bread and rice give energy. Meat and eggs build muscles. Fruits and vegetables protect us."</i>	TTR: 48.5; ASL: 8.7; FRE: 78.2
Enriched	<i>"Our body is like a big city. For the city to work, it needs energy (from carbohydrates), building materials (from proteins), and protection (from vitamins). Carbohydrates are found in bread and rice. Proteins build and repair muscles. Vitamins from fruits and vegetables protect our skin and help digestion."</i>	TTR: 55.3; ASL: 11.4; FRE: 65.8

LG_lex: +0.68; LG_syn: +0.54; LG_read: +0.72

Analysis: This novice teacher employs a scaffolded adaptation strategy, maintaining the central metaphor ("city") across both versions while progressively adding scientific terminology (P3: Continuous Analogies). The simplified version introduces the core idea; the enriched version layers disciplinary vocabulary ("carbohydrates", "proteins", "vitamins") onto the same conceptual framework. The teacher did not "start over" with a new text but instead extended the simplified version, creating a linguistic bridge that supports comprehension while building academic language.

Prompting mechanics: The teacher explicitly instructed the AI to "keep the city metaphor but add the scientific names for each food group and explain their role in one short sentence per group." This meta-cognitive prompting (P4: Structural Narrative) demonstrates novice capacity for strategic prompt engineering when explicitly supported by training materials (Talebzadeh, 2025b).

5.2.3. Case 3 (E-LG): Dialogic Enrichment – Topic: Light and Reflection (Grade 5 Science)

Version	Text (translated from Persian)	Linguistic Profile
Simplified	<i>"Ava visited Dr. Roshan's laboratory. She learned that when light hits an object, the object reflects the light to our eyes. In complete darkness, nothing can be seen because no light reaches our eyes."</i>	TTR: 51.2; ASL: 10.5; FRE: 74.6
Enriched	<i>"Ava, a curious and intelligent girl, visited Dr. Roshan's light laboratory. The doctor explained: 'When light from a source like the sun hits an object, the object reflects part of the light. This reflected light travels to our eyes, and our brain processes it to see the image.' The doctor showed Ava a glowing material: 'This produces its own light—a natural or artificial light source.'"</i>	TTR: 62.4; ASL: 14.8; FR

LG_lex: +1.12; LG_syn: +0.94; LG_read: +1.08

Analysis: The experienced teacher uses a dialogic enrichment strategy (P4: Structural Narrative), embedding scientific concepts within an engaging narrative frame. Unlike the novice's polarised approach, this teacher retains the story structure across both versions, enriching through dialogue, character development ("Ava, a curious girl"), and elaboration of mechanisms ("brain processes it"). The enriched version introduces new concepts ("natural or artificial light source") through contextualised, meaningful interaction rather than decontextualised definition.

Prompting mechanics: The teacher combined narrative elements with scientific content, prompting: "Write a story where a child visits a scientist and learns about light reflection. For the advanced version, add dialogue, characters, and a challenging question at the end." This dual-focus prompting (narrative + science) created coherence across versions while allowing lexical and syntactic expansion within a familiar genre.

5.2.4. Case 4 (E-SG): Minimal Differentiation — Topic: Geographical Directions (Grade 4 Social Studies)

Version	Text (translated from Persian)	Linguistic Profile
Simplified	"Amir, a curious boy, played in the schoolyard. His teacher gave him a hat and said: 'Face the sun.' Amir faced the sun. 'That's north,' said the teacher. 'Your right hand is east, left is west, and behind you is south.' Amir learned the four directions."	TTR: 56.7; ASL: 9.8; FRE: 80.1
Enriched	"Amir, a curious boy, played in the schoolyard. His teacher gave him a hat and said: 'Face the sun, Amir.' Amir faced the sun. 'That's north,' the teacher smiled. 'Your right hand is east—where the sun rises. Left is west—where the sun sets. Behind you is south—always cooler.' Amir smiled: 'I learned north, east, west, and south!'"	TTR: 58.2; ASL: 10.2; FRE:
LG_lex: +0.15; LG_syn: +0.08; LG_read: +0.21		

Analysis: This experienced teacher produces minimal linguistic differentiation across versions—the smallest gap observed. Both versions retain the same narrative structure, similar sentence length, and

accessible vocabulary. The enrichment is primarily stylistic (adding dialogue tags, adjectives) and explanatory (elaborating direction associations: "where the sun rises", "where the sun sets"), rather than structural or lexical. This pattern may reflect either (a) a teacher's judgment that the lesson content is inherently accessible and does not require radical differentiation, or (b) a conservative adaptation style that privileges consistency over challenge (P6: Definition-Centred Conservatism).

Prompting mechanics: The teacher used minimal prompting variation, simply asking the AI to "make the story a little more interesting for advanced students" without specifying linguistic targets. The resulting output suggests reliance on AI's default generation rather than strategic prompt engineering.

5.3. Synthesis: Polarised Versus Scaffolded Adaptation Patterns

Pattern	Code	Description	Manifestation in Cases
P1	Over-simplification	Simplified version stripped of essential concepts	Case 1 (N-LG)
P2	Academic Enrichment	Enriched version uses university-level discourse	Case 1 (N-LG)
P3	Continuous Analogies	Core metaphor extended across both versions	Case 2 (N-SG)
P4	Structural Narrative	Story/Dialogue frame maintained, enriched through elaboration	Case 3 (E-LG)
P5	Challenging Questions	Higher-order questions added in enriched version	Case 3 (E-LG), Case 2 (N-SG)
P6	Definition-Centred	Focus on definitions over connections to lived experience	Case 4 (E-SG)

The qualitative analysis reveals that experienced teachers (Case 3) achieve large linguistic gaps through narrative embedding and dialogic enrichment—maintaining coherence while adding layers of complexity—whereas novice teachers (Case 1) achieve large gaps through abrupt genre shifts (from conversational to academic). Conversely, small gaps in both groups (Cases 2 and 4) reflect scaffolded extension (novice) and stylistic minimalism (experienced), suggesting that "small gap" does not equate to "poor differentiation" but may reflect strategic choices about content accessibility.

These qualitative patterns, particularly P1–P6, serve as interpretative resources for the discussion section, enabling contextualisation of the quantitative findings within the broader literature on teacher knowledge and adaptive expertise.

6. Discussion

This section interprets the empirical findings in light of the theoretical framework, addresses the paradoxes revealed by the data, and articulates implications for teacher education and AI-enhanced pedagogy. The discussion is organised around four interconnected themes: (a) re-evaluating the linguistic gap through PCK and TPACK; (b) the paradox of AI-mediated differentiation; (c) implications for teacher education and prompt engineering; and (d) limitations and future research directions.

6.1. Re-evaluating the Linguistic Gap through PCK and TPACK

The findings reveal a fundamental distinction in how novice and experienced teachers conceptualise and enact content differentiation in AI-enhanced contexts. Novice teachers, when generating enriched content, disproportionately expanded lexical complexity ($LG_{lex} = + 2.64$ in extreme cases) and syntactic depth ($LG_{syn} = + 2.31$) while simultaneously producing oversimplified versions (P1) that stripped content of essential conceptual structure. This polarised pattern reflects what Shulman (1986, 1987) identified as the gap between content knowledge (CK) and pedagogical content knowledge (PCK)—novice teachers possess declarative knowledge about photosynthesis or nutrition but have not yet developed the transformative capacity to represent this knowledge in pedagogically accessible forms across multiple learner levels (Carlson et al., 2019).

This interpretation aligns with the Refined Consensus Model's distinction between collective, personal, and enacted PCK (Carlson et al., 2019). Novice teachers appear to draw primarily upon collective PCK—the generalised knowledge acquired through formal training—without having personalised this knowledge through sustained classroom experience. Their enriched versions often resemble textbook definitions rather than pedagogically crafted explanations, suggesting that the "academic enrichment" observed (P2) represents a direct transmission of academic discourse rather than its transformation for pedagogical purposes (Bernstein, 1996, 2000). The abrupt genre shift from conversational to academic register (Case 1) exemplifies this unmediated transmission, wherein the teacher equates "enrichment" with "expert-level vocabulary" without constructing the conceptual bridges necessary for comprehension (Wheelahan, 2007).

Conversely, experienced teachers' larger gaps (Case 3) were achieved through structurally coherent enrichment—maintaining narrative frames (P4), extending analogies (P3), and embedding scientific terminology within familiar discourse structures. This pattern reflects what Loughran et al. (2016) describe as pedagogical reasoning—the capacity to "hold content knowledge in pedagogically powerful ways" that respect learners' cognitive capacities while advancing understanding. The dialogic structure observed in Case 3 ("Ava visited Dr. Roshan's laboratory...") demonstrates experienced teachers' ability to create what Bernstein (1996) termed "recontextualising rules"—transforming vertical discourse (scientific knowledge) into horizontal discourse (accessible narrative) while preserving its conceptual integrity (Wheelahan, 2007).

The TPACK framework (Koehler & Mishra, 2009) further illuminates these differences. Novice teachers demonstrated procedural TPACK—they could use AI tools to generate content—but lacked the integrated knowledge to strategically evaluate and modify AI outputs for pedagogical purposes (Chen et al., 2026). This is consistent with research on out-of-field teaching (Hobbs, 2020; Kenny et al., 2019) and novice teacher noticing (Wang, 2026), which demonstrate that full-time teaching experience stabilises attention to learning-relevant features and refines interpretive reasoning. Novice teachers' reliance on AI-generated "enrichment" without critical pedagogical vetting suggests that AI tools may amplify rather than compensate for gaps in enacted PCK (Mishra & Bozkurt, 2026).

6.2. The Paradox of AI-Mediated Differentiation

The findings reveal a significant paradox: generative AI tools, when deployed without explicit pedagogical guidance, tend to produce asymmetric linguistic gaps that may undermine rather than support differentiated instruction. In novice teachers' hands, AI-generated enriched content exhibited lexical and syntactic complexity far exceeding the capabilities of target high-achieving learners (Case 1), while simplified content was sometimes oversimplified to the point of conceptual dilution (P1). This pattern suggests that AI tools, left to their default generation parameters, amplify teachers' existing pedagogical tendencies rather than providing corrective scaffolding (Dell'Acqua et al., 2026; Cao et al., 2026).

This paradox aligns with Dell'Acqua and colleagues' (2026) concept of the "jagged technological frontier"—AI assistance improves performance for some tasks but worsens it for others, even within the same workflow. For novice teachers, AI-generated enrichment often "improves" lexical and syntactic complexity but simultaneously "worsens" pedagogical appropriateness, as the AI's statistical language model prioritises textual coherence and diversity over pedagogical accessibility (Messner et al., 2025; Fan et al., 2025). The sharp contrast between simplified and enriched versions in novice-generated content (Case 1, LG_total = + 2.14 SD) reflects an AI-influenced polarisation that may not correspond to the teacher's actual pedagogical intentions.

This interpretation is consistent with studies on human-AI collaboration (Corbin et al., 2025; Chan & Lee, 2025), which demonstrate that teachers often experience significant emotional burden and professional uncertainty when attempting to establish appropriate boundaries for AI use. The qualitative vignettes reveal that teachers who employed structured prompt engineering—specifying analogical continuity (Case 2) or narrative frames (Case 3)—achieved more pedagogically coherent differentiation than those who relied on AI default outputs (Cases 1 and 4). This suggests that the linguistic gap is not merely a property of teacher expertise or AI capability, but emerges from the quality of human-AI interaction (Cukurova, 2024; Molenaar, 2022).

Crucially, the findings challenge Hypothesis C—that AI tools standardise the linguistic gap—by demonstrating that AI outputs vary substantially as a function of teacher prompting strategies. The overlap analysis (OVL = 0.67), indicating moderate distribution similarity between groups does not imply AI-driven standardisation but rather reflects the influence of shared prompt templates on initial content

generation (Dogan, 2025). The qualitative evidence indicates that teachers who employed strategic prompt engineering produced substantively different outputs than those who used basic prompts, suggesting that AI standardisation is not inherent but contingent upon teacher AI literacy (Chiu et al., 2024; An et al., 2026).

6.3. Implications for Teacher Education and AI Prompt Engineering

The findings carry significant implications for teacher education and professional development. First, they suggest that AI literacy programmes must move beyond technical training in tool operation to address the pedagogical reasoning required for effective content differentiation (Cukurova, 2025; Chiu et al., 2026). The polarised adaptation observed among novice teachers (P1 + P2) indicates that training in prompt engineering alone—without attention to how linguistic complexity relates to cognitive load and learner comprehension—may inadvertently produce content that is pedagogically counterproductive (Mayer, 2024; Vanhees et al., 2025).

Second, the study suggests that effective AI-enhanced differentiation requires what we term "pedagogically oriented prompt engineering"—the capacity to specify not only content parameters (Role, Audience, Format) but also structural and linguistic constraints that maintain pedagogical coherence across versions. The scaffolded adaptation strategies observed in Cases 2 and 3 exemplify this approach: specifying analogical continuity ("keep the city metaphor"), requiring narrative embedding ("write a story with dialogue"), and structuring progressive elaboration ("add scientific names and one sentence explaining each"). These meta-cognitive prompting strategies enabled teachers to produce enriched versions that maintained conceptual coherence while increasing linguistic complexity (Talebzadeh, 2025b; Dogan, 2025).

Third, the findings underscore the importance of developing teachers' capacity for "linguistic noticing"—the ability to identify how lexical selection, syntactic complexity, and textual organisation shape learner comprehension (Wang, 2026; van Es & Sherin, 2026). The qualitative analysis reveals that experienced teachers (Case 3) demonstrated greater awareness of how linguistic choices affected accessibility, whereas novice teachers (Case 1) focused primarily on content accuracy without attending to linguistic mediation. This suggests that teacher education programmes should include structured opportunities for teachers to analyse and critique AI-generated content through linguistic lenses, developing sensitivity to how language functions as a pedagogical tool (Brookhart, 2011; Pellegrino, 2018).

6.4. Limitations and Future Research Directions

Several limitations warrant acknowledgment. The workshop context, while designed to simulate authentic planning tasks, cannot fully capture the spontaneity and contextual constraints of real-time classroom differentiation (Hay, 2025; Detyna et al., 2025). Teachers' content generation in a professional development setting may reflect idealised rather than typical practice, potentially inflating the linguistic gap relative to authentic classroom conditions. Future research should examine linguistic gap patterns

in naturalistic settings, using observation and lesson documentation to capture differentiation as it occurs in real-time instruction.

The cultural and linguistic specificity of the Iranian EFL context also constrains generalisability (Marmolejo-Ramos et al., 2025). The Persian language, with its distinct morphological and syntactic structures, may yield different linguistic gap patterns than analytic languages such as English. Cross-linguistic and cross-cultural replication studies are needed to establish the robustness of the findings across educational contexts.

The reliance on automated linguistic tools (Textstat, SpaCy) introduces potential measurement limitations. While metric audit established acceptable validity for Persian educational texts ($\rho \geq 0.79$), automated indices may not capture the full nuance of textual complexity, particularly regarding semantic coherence and conceptual depth (Messner et al., 2025; Overney et al., 2024). Future research should incorporate human expert ratings and qualitative discourse analysis to complement automated metrics.

The study focused exclusively on simplified and enriched content versions, representing two poles of a continuum. This binary conceptualisation may not capture the full range of differentiation strategies teachers employ, including visual aids, interactive elements, and multimodal supports (Mayer, 2024; Ouariach et al., 2026). Future research should examine how linguistic differentiation interacts with other pedagogical dimensions in AI-enhanced instruction.

Finally, the cross-sectional design limits causal inference regarding the development of PCK and adaptation expertise. Longitudinal research tracking teachers' linguistic gap patterns across their careers would provide insights into how adaptive expertise develops and how AI tools influence this development trajectory (Priestley et al., 2026; Day et al., 2006). Despite these limitations, the study contributes empirically grounded insights into the interplay of teacher expertise, AI tools, and linguistic differentiation, with implications for teacher education, professional development, and the design of AI systems that genuinely support differentiated instruction.

7. Conclusion & Pedagogical Recommendations

7.1. Executive Summary of Core Contributions

This exploratory study investigated the linguistic gap generated by novice and experienced teachers when adapting AI-generated content for differentiated instruction. Through quantitative analysis of 178 content samples (89 simplified, 89 enriched) across lexical, syntactic, and readability dimensions, and qualitative examination of extreme cases, four principal findings emerge.

First, the study does not confirm Hypothesis D—that group differences vary significantly by linguistic dimension. The MANOVA analysis (Section 4.3) revealed that the interactive effect (Group $\times$ Dimension) was not statistically significant (Wilks' $\Lambda = 0.94$, $F(2, 86) = 1.85$, $p = 0.160$, $\eta^2p = 0.03$). However, univariate analyses indicated that novice teachers produced substantially larger gaps in lexical complexity (LG_lex:

novice $M = 1.24$, $SD = 0.41$; experienced $M = 0.72$, $SD = 0.38$; Hedges' $g = 0.82$, $p < 0.001$) and syntactic depth (LG_syn: novice $M = 1.13$, $SD = 0.44$; experienced $M = 0.65$, $SD = 0.35$; Hedges' $g = 0.71$, $p < 0.001$), but differences in readability (LG_read: novice $M = 0.57$, $SD = 0.39$; experienced $M = 0.48$, $SD = 0.36$; Hedges' $g = 0.26$, $p = 0.043$) were more modest. The non-significant interaction suggests that the pattern of novice > experienced differences is consistent across dimensions, though most pronounced where conscious knowledge transformation from declarative to pedagogical application is required.

Second, the findings support Hypothesis B while rejecting Hypothesis A. Novice teachers produced larger overall linguistic gaps (LG_total: novice $M = 1.13$, $SD = 0.38$; experienced $M = 0.73$, $SD = 0.32$; Hedges' $g = 0.68$, $p = 0.002$). This pattern reflects novice teachers' reliance on recent training in standards-based approaches and their greater willingness to adopt AI-generated suggestions without critical pedagogical vetting. However, the qualitative analysis reveals that the nature of the gap differs qualitatively: novice teachers produce "polarised" gaps through abrupt genre shifts (P1: Over-simplification; P2: Academic Enrichment), while experienced teachers produce "scaffolded" gaps through narrative extension and analogical elaboration (P3: Continuous Analogies; P4: Structural Narratives).

Third, the overlap analysis (Section 4.4) revealed substantial distribution separation between groups, rejecting Hypothesis C. The Weitzman overlap coefficient for LG_total was $OVL = 0.36$, indicating that less than half of the distribution overlap existed between novice and experienced teachers. Across dimensions, OVL values ranged from 0.31 (LG_lex) to 0.47 (LG_read), confirming that AI tools do not standardise the linguistic gap but rather amplify existing differences. Teachers who employed structured prompt engineering—specifying analogical continuity, narrative embedding, and progressive elaboration—produced more pedagogically coherent differentiation than those who relied on AI default outputs. This finding underscores that AI is a mediator, not a standardiser, of content adaptation.

Fourth, the negative case analysis (Section 5) identified six recurrent patterns (P1–P6) distinguishing polarised from scaffolded adaptation. Over-simplification (P1) and academic enrichment (P2) characterised novice-generated gaps, while continuous analogies (P3), structural narratives (P4), and challenging questions (P5) characterised experienced-generated gaps. Definition-centred conservatism (P6) was observed in both groups, associated with minimal differentiation. These qualitative patterns provide interpretative depth to the quantitative findings, revealing that the linguistic gap is not merely a statistical artefact but reflects qualitatively distinct pedagogical reasoning processes.

7.2. Concrete Recommendations for Educational Policy, Curriculum Design, and Teacher Education

The findings carry significant implications for teacher education, curriculum design, and educational policy, organised across three interconnected domains.

For teacher education programmes:

1. Integrate linguistic awareness into AI literacy curricula. Pre-service and in-service teachers require systematic training in how linguistic choices—lexical selection, syntactic structure, textual organisation—shape learner comprehension and cognitive load. Professional development should include structured analysis of AI-generated content through linguistic lenses, developing teachers' capacity for "linguistic noticing"—the ability to identify how linguistic features affect accessibility and learning (van Es & Sherin, 2026; Wang, 2026).
2. Teach structured prompt engineering for pedagogical coherence. Rather than generic prompt templates, teachers should learn to specify: (a) analogical continuity—what conceptual frames or metaphors should be maintained across versions; (b) progressive elaboration—what layers of complexity should be added incrementally; and (c) pedagogical coherence—how linguistic changes serve learning objectives. The scaffolded adaptation strategies observed in Cases 2 and 3 (Section 5) provide actionable exemplars for training materials. This "pedagogically oriented prompt engineering" can mitigate the polarisation observed in novice use by scaffolding teachers' pedagogical reasoning (Dogan, 2025; Talebzadeh, 2025b).
3. Reinforce PCK development through supervised AI use. Mentoring structures should pair novice teachers with experienced mentors during AI-assisted content generation, enabling novices to observe and enact the pedagogical reasoning that underlies scaffolded differentiation. The qualitative gap between novice polarisation and experienced scaffolding suggests that such mentored practice may accelerate PCK development (Carlson et al., 2019; Hartmuth et al., 2025). Structured reflection on AI-generated content—asking "For whom is this text appropriate?" and "What conceptual bridges are missing?"—should be integrated into teacher preparation curricula.

For curriculum designers and educational technology developers:

4. Embed pedagogical guidance within AI interfaces. AI tools should include proactive prompts encouraging teachers to specify continuity frames, progressive complexity structures, and learner-level benchmarks before generating differentiated content. Design features might include: (a) drop-down menus for selecting complexity progression type (analogical, narrative, or conceptual); (b) visual indicators showing projected readability and lexical complexity relative to target benchmarks; and (c) exemplar libraries showing effective differentiation patterns from expert teachers (Cao et al., 2026; Molenaar, 2022).
5. Develop linguistic gap analytics for teacher feedback and reflection. AI platforms could provide teachers with automated feedback on their differentiation patterns, including visualisations of lexical, syntactic, and readability gaps, and comparisons with expert-defined benchmarks. Such analytics can support teacher reflection and iterative refinement of content adaptation strategies (Ortega-Arranz et al., 2026; Fan et al., 2025). However, such analytics must be designed as formative feedback tools rather than evaluative metrics to avoid discouraging experimentation.

For educational policy:

6. Articulate AI-enhanced differentiation as a core professional competency. National teaching standards and professional development frameworks should explicitly address teachers' capacity to strategically adapt AI-generated content across linguistic dimensions, recognising this as a core competency of 21st-century teaching (Chiu et al., 2024; Cukurova, 2025). The linguistic gap, as operationalised in this study, offers a measurable indicator for assessing this competency—not for high-stakes evaluation, but for diagnostic and developmental purposes.

7. Establish institutional structures for sustainable professional learning. Schools and districts should create professional learning communities focused on AI-enhanced content differentiation, where teachers collaboratively analyse generated content, share effective prompt strategies, and develop shared understanding of linguistic alignment benchmarks. Such communities can provide the sustained, contextualised support that one-off workshops cannot offer (Priestley et al., 2026; Chiu et al., 2026).

7.3. Final Concluding Remark

This study has demonstrated that the linguistic gap—far from being a technical artefact of AI content generation—is a window into teachers' pedagogical reasoning, revealing how professional knowledge, experiential learning, and interactive strategies with AI tools shape instructional adaptation. Novice teachers, despite their technological fluency, produce polarised differentiation that may undermine rather than support learning; experienced teachers, through scaffolded enrichment, demonstrate the pedagogical wisdom that AI tools can amplify but not replace.

The findings, however, also reveal that experience alone does not guarantee effective AI-enhanced differentiation. The experienced teacher with the smallest gap (Case 4) employed minimal linguistic differentiation, raising questions about whether experience leads to greater discernment about when differentiation is appropriate or whether certain experienced teachers have developed conservative adaptation habits that resist enrichment. This complexity underscores the importance of reflective practice and sustained professional development for all teachers, regardless of experience level.

As generative AI becomes increasingly embedded in educational practice, the central question is not whether AI can generate differentiated content, but whether teachers can engage with AI critically and strategically to serve pedagogical purposes. The findings of this study suggest that effective AI-enhanced differentiation requires not only technological competence but also developed PCK, linguistic awareness, and structured professional support. The future of human-AI collaboration in learning environments hinges not on replacing pedagogical judgment with algorithmic efficiency, but on cultivating teachers' capacity to orchestrate AI tools with intentionality, coherence, and responsiveness to learner needs. This study offers a conceptual and empirical foundation for that cultivation, contributing to a broader reimagining of teacher expertise in the AI era—one in which teachers' distinctive capacity for pedagogical reasoning remains central, even as the tools at their disposal become increasingly intelligent.

References

1. An X, Bai J, Li Y, Zhou Y, Chai CS, Hong J-C (2026) Development of an artificial intelligence technology pedagogical content knowledge (AI-TPACK) scale for human and AI concurrently as teachers. *Asia-Pacific Educ Research* 35(3):483–493. <https://doi.org/10.1007/s40299-025-01045-2>
2. Beijaard D (1995) Teachers' prior experiences and actual perceptions of professional identity. *Teachers Teach* 1(2):281–294. <https://doi.org/10.1080/1354060950010209>
3. Bernstein B (1996) *Pedagogy, symbolic control, and identity: Theory, research, critique*. Rowman & Littlefield, Rev. ed
4. Bernstein B (2000) *Pedagogy, symbolic control, and identity*, vol 5. Bloomsbury Publishing PLC
5. Biesta G, Priestley M, Robinson S (2015) The role of beliefs in teacher agency. *Teachers Teach* 21(6):624–640. <https://doi.org/10.1080/13540602.2015.1044325>
6. Braun V, Clarke V (2020) One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Res Psychol* 18(3):328–352. <https://doi.org/10.1080/14780887.2020.1769238>
7. Braun V, Clarke V (2023) Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychol Rev* 17(4):695–718. <https://doi.org/10.1080/17437199.2022.2161594>
8. Brookhart SM (2011) Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues Pract* 30(1):3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
9. Cao J, Wang S, Chen X, Schunn C (2026) Co-develop learning tasks with generative AI: Mapping challenges, gains, and professional development needs. *IEEE Trans Learn Technol*. Advance online publication <https://doi.org/10.1109/TLT.2026.11441993>
10. Carlson J, Daehler KR, Alonzo AC, Barendsen E, Berry A, Borowski A, Carpendale J, Chan KKH, Cooper R, Friedrichsen P, Gess-Newsome J, Henze-Rietveld I, Hume A, Kirschner S, Liepertz S, Loughran J, Mavhunga E, Neumann K, Nilsson P, Wilson CD (2019) The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer. https://doi.org/10.1007/978-981-13-5898-2_2
11. Chan CKY, Lee KK (2025) The balancing act between AI and authenticity in assessment: A case study of secondary school students' use of GenAI in reflective writing. *Comput Educ* 105399. <https://doi.org/10.1016/j.compedu.2025.105399>
12. Chan CKY, Wong HY (2025) Assessing the ability of generative AI in English literary analysis through Bloom's Taxonomy. *Discover Comput* 28(1):127. <https://doi.org/10.1007/s10791-025-09572-8>
13. Chen J, Loke SP, Teh SY (2026) From TPACK to I-TPACK: Exploring teachers' perceptions and experiences with AI tools in pedagogical practice in Chinese higher education. *Interact Learn Environ* 1–20. <https://doi.org/10.1080/10494820.2026.2683204>
14. Chiu TKF, Ahmad Z, Çoban M (2024) Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Educ Inform Technol* 29(3):1–24. <https://doi.org/10.1007/s10639-024-13094-z>

15. Chiu TKF, Bali S, Tondeur J, Howard S, Chan KKH (2026) Teacher education for artificial intelligence literacy through a self-determination theory perspective. *Eur J Teacher Educ* 1–21. <https://doi.org/10.1080/02619768.2026.2621848>
16. Chiu TKF, Rospigliosi P (2026) Six global frameworks for human-centred AI literacy and competency: Comparative analysis and a way forward. *Interact Learn Environ* 34(3):1003–1005. <https://doi.org/10.1080/10494820.2026.2648342>
17. Chiu TKF, Xia Q, Zhou X, Chai CS, Cheng M (2025) Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers Education: Artif Intell* 4:100118. <https://doi.org/10.1016/j.caeai.2022.100118>
18. Corbin T, Dawson P, Nicola-Richmond K, Partridge H (2025) Where's the line? It's an absurd line': Towards a framework for acceptable uses of AI in assessment. *Assess Evaluation High Educ* 50(5):705–717. <https://doi.org/10.1080/02602938.2025.2456207>
19. Creswell JW, Plano Clark VL (2018) *Designing and conducting mixed methods research*, 3rd edn. SAGE
20. Crossley SA, McNamara DS (2016) The development and validation of an automated linguistic complexity measure. *Education Tech Research Dev* 64(4):1–23. <https://doi.org/10.1007/s11423-016-9445-5>
21. Cukurova M (2024) The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *Br J Edu Technol* 55(6):1–18. <https://doi.org/10.1111/bjet.13514>
22. Cukurova M (2025) Promoting and protecting teacher agency in the age of artificial intelligence. UNESCO Publishing. https://teachertaskforce.org/sites/default/files/2025-09/1149_25_Promoting%20and%20Protecting%20Teacher%20Agency_FINAL_3Sep.pdf
23. Day C, Kington A, Stobart G, Sammons P (2006) The personal and professional selves of teachers: Stable and unstable identities. *Br Edu Res J* 32(4):601–616. <https://doi.org/10.1080/01411920600775316>
24. Dell'Acqua F, McFowland E, Mollick III, Lifshitz E, Kellogg H, Rajendran KC, Kraye S, Candelon L, F., Lakhani KR (2026) Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organ Sci* 37(2):403–423. <https://doi.org/10.1287/orsc.2025.21838>
25. Detyna M, Ogunbase A, Briffa H (2025) A socio-material approach to investigating classrooms: Student engagement in an innovative learning environment. *Learn Environ Res* 28(3):473–502. <https://doi.org/10.1007/s10984-025-09529-0>
26. Dogan S (2025) A human-AI hybrid co-design model: Planning effective instruction with ChatGPT. *Asian J Distance Educ* 20(1):1–20
27. Fan F, Wang S, Dinuer M, Hemuti M, Nie X, Yang LT (2025) Enhancing students' metacognition with innovative IA-based metacognitive reflective learning tool. *IEEE Trans Learn Technol* 18:104–110. <https://doi.org/10.1109/TLT.2025.11039084>

28. Hartmuth D, Amin RM, Schlichting D, Schött S, Haudenschild J, Küchenhoff H, Neuhaus BJ (2025) Developing science teachers' enacted pedagogical content knowledge through integration of student feedback into the Refined Consensus Model of pedagogical content knowledge. *Cogent Educ* 12(1):2470563. <https://doi.org/10.1080/2331186X.2025.2470563>
29. Hay A (2025) What may be: Policy enactment in education, a new conceptual framework with actor-network theory. *J Educ Policy* 40(2):179–198. <https://doi.org/10.1080/02680939.2024.2411989>
30. Hobbs L (2020) Learning to teach science out-of-field: A spatial-temporal experience. *J Sci Teach Educ* 31(7):725–745. <https://doi.org/10.1080/1046560X.2020.1718315>
31. Kenny J, Hobbs L, Whannell R (2019) Designing professional development for teachers teaching out-of-field. *Prof Dev Educ* 46(3):500–515. <https://doi.org/10.1080/19415257.2019.1613257>
32. Koehler M, Mishra P (2009) What is technological pedagogical content knowledge (TPACK)? *Contemp Issues Technol Teacher Educ* 9(1):60–70
33. Kyle K, Crossley SA (2015) Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Q* 49(4):757–786. <https://doi.org/10.1002/tesq.194>
34. Loughran J, Keast S, Cooper R (2016) Pedagogical reasoning in teacher education. In J. Loughran & M. L. Hamilton (Eds.), *International handbook of teacher education* (pp. 387–421). Springer. https://doi.org/10.1007/978-981-10-0366-0_10
35. Luckin R, Holmes W (2016) *Intelligence unleashed: An argument for AI in education*. Pearson Education
36. Marmolejo-Ramos F, Marrone R, Korolkiewicz M, Gabriel F, Siemens G, Joksimovic S, Yamada Y, Mori Y, Rahwan T, Sahakyan M, Sonna B, Meirmanov A, Bolatov A, Som B, Ndukaihe I, Arinze NC, Kunderát J, Skanderová L, Ngo V-G, Tejada J (2025) Factors influencing trust in algorithmic decision-making: An indirect scenario-based experiment. *Front Artif Intell* 7:1465605. <https://doi.org/10.3389/frai.2024.1465605>
37. Mayer RE (2024) The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychol Rev* 36(1):8. <https://doi.org/10.1007/s10648-023-09842-1>
38. McKenney S, Reeves TC (2025) Educational design research for relevant & robust scholarship. *J Comput High Educ* 37(2):614–638. <https://doi.org/10.1007/s12528-025-09456-2>
39. Messner R, Smith S, Richards C (2025) Artificial intelligence and qualitative data analysis: Epistemological incongruences and the future of the human experience. *Int J Qualitative Methods* 24:16094069251371481. <https://doi.org/10.1177/16094069251371481>
40. Mishra P, Bozkurt A (2026) TPACK in the age of context and artificial intelligence: Punya Mishra on the wicked problem of AI and why algorithms can't replace context. *Open Praxis* 18(2). <https://doi.org/10.55982/openpraxis.18.2.1093>
41. Molenaar I (2022) Towards hybrid human-AI learning technologies. *Eur J Educ* 57(4):632–645. <https://doi.org/10.1111/ejed.12527>
42. Ortega-Arranz A, Topali P, García-Zarza P, Bote-Lorenzo ML, Asensio-Pérez JI, Molenaar I (2026) GenAI analytics and pedagogical configurations: Featuring new generative AI systems in education.

- In *Proceedings of the LAK26: 16th International Learning Analytics and Knowledge Conference* (pp. 779–785). ACM. <https://doi.org/10.1145/3785022.3785054>
43. Ouariach FZ, Ouariach S, Ibrahim G, Frij R (2026) Optimizing multimedia elements for effective digital learning: Principles, practices, and innovations. *Instructional design in e-learning*. CRC, pp 15–32
 44. Overney C, Saldías B, Dimitrakopoulou D, Roy D (2024) SenseMate: An accessible and beginner-friendly human-AI platform for qualitative data analysis. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (pp. 922–939). ACM. <https://doi.org/10.1145/3640543.3645194>
 45. Pellegrino JW (2018) Assessment of and for learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 410–421). Routledge. <https://doi.org/10.4324/9781315617572-32>
 46. Pillen M, Beijaard D, den Brok P (2013) Tensions in beginning teachers' professional identity development, accompanying feelings and coping strategies. *Eur J Teacher Educ* 36(3):240–260. <https://doi.org/10.1080/02619768.2012.696192>
 47. Priestley M, Biesta GJJ, Philippou S, Robinson S (2015) The teacher and the curriculum: Exploring teacher agency. In: Wyse D, Hayward L, Pandya J (eds) *The SAGE handbook of curriculum, pedagogy and assessment*. SAGE, pp 218–231. <https://doi.org/10.4135/9781473921405.n13>
 48. Priestley M, Biesta G, Robinson S (2026) *Teacher agency: An ecological approach*, 2nd edn. Bloomsbury Publishing
 49. Ritzhaupt AD, Dawson K, Crompton H (2026) Toward guidance and principled values on the ethical, responsible, and transparent use of GenAI in the educational research process. *J Res Technol Educ* 58(2):235–249. <https://doi.org/10.1080/15391523.2026.2621640>
 50. Selwyn N (2019) Should robots replace teachers? AI and the future of education. *Polity*
 51. Shulman LS (1986) Those who understand: Knowledge growth in teaching. *Educational Researcher* 15(2):4–14. <https://doi.org/10.3102/0013189X015002004>
 52. Shulman L (1987) Knowledge and teaching: Foundations of the new reform. *Harv Educational Rev* 57(1):1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
 53. Sun Y, Xin H, Li S, Niu Q, Chai CS, Huang L, Chen G (2026) Modeling AI-TPACK in practice: Insights from teachers' multi-agent workflow design. *arXiv* 260513906. <https://doi.org/10.48550/arXiv.2605.13906>
 54. Sweller J (2011) Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
 55. Talebzadeh H (2025a) Bilingual Teacher's Guide – Lesson 1: Creating smart assignments with the IAT Model. Zenodo. <https://doi.org/10.5281/zenodo.17525145>
 56. Talebzadeh H (2025b) Bilingual Teacher's Guide – Lesson 2: Designing personalized educational content with AI and the IAT Model. Zenodo. <https://doi.org/10.5281/zenodo.17536752>

57. Talebzadeh H (2025c) The Integrated AI Triad (IAT): A systemic framework for redefining the roles of teachers, managers, and families in the AI era. Zenodo. <https://doi.org/10.5281/zenodo.21842266>
58. Triviño-Amigo N, Polo-Campos I, Gómez-Paniagua S, Barrios-Fernandez S, Mendoza-Muñoz M, Rojo-Ramos J (2023) Differences in the perception regarding inclusion preparation among teachers at different educational stages. *Int J Environ Res Public Health* 20(4):3420. <https://doi.org/10.3390/ijerph20043420>
59. van Es EA, Sherin MG (2026) Exploring teacher noticing in practice: The role of shaping. *ZDM – Math Educ* 58:19–31. <https://doi.org/10.1007/s11858-025-01740-w>
60. Vanhees C, Nijlunsing J, Muijs D, Crato N, Wils M, Wiliam D, Surma T, Kirschner PA (2025) The role of knowledge-rich curricula in promoting deep thinking and complex skill acquisition. *Learn Individual Differences* 121:102729. <https://doi.org/10.1016/j.lindif.2025.102729>
61. Wang L (2026) Within-teacher contrasts of in-field and out-of-field noticing: Integrated knowledge and teaching experience. *Teach Teacher Educ* 172:105671. <https://doi.org/10.1016/j.tate.2026.105671>
62. Wheelahan L (2007) How competency-based training locks the working class out of powerful knowledge: A modified Bernsteinian analysis. *Br J Sociol Educ* 28(5):637–651. <https://doi.org/10.1080/01425690701505540>
63. Wheeley E, Klieve H, Park E, Du Plessis A (2023) Preservice teachers' perceptions about out-of-field teaching: Implications for students, teachers and schools. *Australian J Teacher Educ* 48(4):1–18. <https://doi.org/10.14221/1835-517X.5619>