Conceptual Article

EDUCATIONAL POINT
An International Peer-Reviewed Journal

https://www.educationalpoint.net
ISSN: 3062-1011 (Online)

Intelligent Pronunciation Learning Pathways in Education 4.0: An Instructional Design Framework

Rizgar Qasim Mahmood [1*]

[1] Salahaddin University-Erbil, IRAQ
* Correspondence: rizgar.mahmood@su.edu.krd

CITATION: Mahmood, R. Q. (2026). Intelligent Pronunciation Learning Pathways in Education 4.0: An Instructional Design Framework. *Educational Point*, *3*(3), e191. https://doi.org/10.71176/edup/20013

ARTICLE INFO

Received: 11 April 2026
Accepted: 25 September 2026

OPEN ACCESS

ABSTRACT

Pronunciation plays a central role in oral intelligibility, yet many technology-enhanced language courses still treat it as an add-on or reduce it to isolated drill practice. This conceptual study proposes an instructional design framework for intelligent pronunciation learning pathways within Education 4.0 ecosystems. The study uses a theory-synthesis approach: peer-reviewed work on pronunciation instruction, high-variability phonetic training (HVPT), corrective feedback (CF), computer-assisted pronunciation training (CAPT), automatic speech recognition (ASR), adaptive learning, and learning analytics ethics was selected, compared, and synthesized to identify design requirements for data-informed pronunciation support. The framework contains three interdependent layers: Input, which organizes high-variability exposure through HVPT principles; Feedback, which uses AI-mediated support informed by L2 CF and mediation theory; and Adaptation, which sequences learner-specific pathways through learning analytics and adaptive learning concepts. The framework differs from single-tool CAPT models by treating pronunciation learning as a governed pathway that connects input design, feedback reliability, teacher mediation, contextual constraints, and ethical safeguards. The article also examines Iraqi Kurdistan as an illustrative under-resourced EFL context, where limitations in digital infrastructure, teacher preparation, and pronunciation resources shape implementation. Since the study does not report empirical data, the framework should be read as a design proposition that requires future validation through classroom pilots, design-based research, and fairness audits.

Keywords: Education 4.0, pronunciation, HVPT, corrective feedback, ASR/CAPT, adaptive learning

INTRODUCTION

Education 4.0 has been described as a systemic educational response to the technological and organizational shifts associated with the Fourth Industrial Revolution. Recent syntheses emphasize that Education 4.0 is not reducible to “new tools” but entails integrated configurations of technologies, pedagogies, and governance arrangements that enable personalization, data-informed decision-making, hybrid learning spaces, and

Copyright © 2026 by the authors; licensee: Educational Point. This open-access article is distributed under the terms and conditions of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).

scalable institutional architectures (Souza & Debs, 2024). In this orientation, the unit of analysis is increasingly the learning ecosystem, a network of platforms, stakeholders, data flows, and instructional designs, rather than a single application (Chituc, 2021). Within language education, however, ecosystem-level design has often progressed unevenly across skill domains. For example, reading and writing have long benefited from mature forms of analytics (e.g., automated scoring, corpus-informed feedback), while speaking and pronunciation, particularly at scale, have historically remained difficult to individualize because they demand high-frequency feedback, stable models, and sustained practice. Yet this historical constraint is changing. Research in computer-assisted pronunciation training (CAPT) and automatic speech recognition (ASR) now documents measurable benefits for pronunciation development, and meta-analytic work suggests that ASR-supported interventions can yield medium-sized improvements overall, especially when feedback is explicit and instructionally aligned (Almusharraf et al., 2024; Hirschi et al., 2025; Liu et al., 2025; Thi-Nhu Ngo et al., 2024).

At the same time, speech technologies do not solve pronunciation pedagogy on their own. Systematic reviews of CAPT research show that many interventions still rely on listen-and-repeat or read-aloud drills, privilege segmental accuracy, and underrepresent suprasegmentals, intelligibility, and communicative transfer (Amrate & Tsai, 2025; Crowther & Loewen, 2025; Liu et al., 2025). This design gap matters because contemporary pronunciation pedagogy distinguishes accentedness from comprehensibility and intelligibility and therefore resists native-like accent as the default endpoint (Saito, 2021b). A second gap concerns governance. Education 4.0 ecosystems depend on learner data, yet learning analytics research has warned that data-rich systems can intensify problems of privacy, consent, surveillance, and unequal consequences (Slade & Prinsloo, 2013).

In pronunciation learning, this risk becomes more serious because ASR performance can vary across speech communities, dialects, and accents, which may lead automated systems to penalize or misdiagnose some learners (Koenecke et al., 2020). This manuscript addresses these pedagogical and ethical gaps by proposing an instructional design framework for intelligent pronunciation learning pathways within Education 4.0 ecosystems. The framework integrates:

1. Input designed through high-variability phonetic training (HVPT) principles to support robust perceptual category formation and generalization.
2. Feedback delivered through AI mediation that operationalizes a continuum from implicit signals to explicit correction and graduated scaffolding.
3. Adaptation through learner modeling and sequencing rules that personalize progression while accounting for uncertainty, engagement, and contextual constraints (Thi-Nhu Ngo et al., 2024).

Before future empirical studies can test intelligent pronunciation learning pathways, the field first needs a framework that defines what such pathways should contain, how their components interact, and how researchers can evaluate them across unequal educational settings. This article, therefore, develops a conceptual theory synthesis. It brings together pronunciation pedagogy, speech technology research, adaptive learning, and ethics to develop design architecture for subsequent implementation, comparison, and validation. Iraqi Kurdistan serves as an illustrative application case rather than a data source.

The context matters because research has documented teachers' digital knowledge gaps, uneven infrastructure, and limited pronunciation resources, which test whether an Education 4.0 framework can operate beyond high-resource settings (Ali & Mohammadzadeh, 2022; Mahmood, 2024b).

METHODOLOGICAL APPROACH FOR THE CONCEPTUAL THEORY SYNTHESIS

This article uses conceptual theory synthesis because intelligent pronunciation pathways need conceptual groundwork before they can be tested empirically. Future intervention studies will need more than a new tool or a set of classroom activities; they will need clear constructs, design relations, and evaluative boundaries. The present article, therefore, builds the architecture that later empirical work can validate what counts as structured input, how AI-mediated feedback should guide learner attention, how adaptation should sequence learning, and what ethical limits should govern learner data. In this sense, the article treats conceptual work as a necessary design stage, not as a substitute for empirical evidence. Following Jaakkola (2020), the reviewed literature was read for its theory-building value and for the design problems it helped solve.

Search Strategy and Study Selection

The literature search was designed to identify studies that could justify the need for the Input-Feedback-Adaptation framework (See the **Appendix** for the studies). It was not designed as a systematic review, bibliometric analysis, or PRISMA-style mapping exercise. Instead, it created a narrow and replicable evidence trail for selecting studies that could inform the construction of the framework.

The same keyword logic was used across Web of Science and Scopus, with only the database-specific field codes adjusted. In Web of Science, the search was run in the Topic field for the period 2010–2026 and returned 12 records; all 12 were read for relevance to the framework. In Scopus, the search was run in TITLE-ABS-KEY and limited to the Social Sciences and Arts and Humanities subject areas; it returned 11 records, all of which were read. The master keyword string was: (“second language pronunciation” OR “L2 pronunciation” OR “EFL pronunciation” OR “ESL pronunciation”) AND (“computer-assisted pronunciation training” OR CAPT OR “automatic speech recognition” OR ASR OR “AI-mediated feedback” OR “AI feedback”) AND (“high variability phonetic training” OR HVPT OR “corrective feedback” OR “adaptive learning” OR “learning analytics” OR “personalized learning” OR “Education 4.0”). The search was intentionally narrow because the purpose was not to produce a systematic map of CAPT, ASR, HVPT, or adaptive learning research, but to identify studies that could inform the conceptual design of the Input–Feedback–Adaptation framework. After duplicate records across Web of Science and Scopus were removed, the final set comprised 18 unique studies, all of which were read in full and used to inform the conceptual synthesis and development of the framework.

Inclusion Criteria

Studies were included when they addressed at least one framework-building concern: L2 pronunciation instruction or assessment; HVPT and perceptual input; CAPT, ASR, or AI-supported pronunciation learning; CF, scaffolding, or sociocultural mediation; adaptive learning, learner modeling, or learning analytics; or ethical and contextual constraints in data-driven language education. Peer-reviewed articles, systematic reviews, meta-analyses, and theoretically relevant book chapters formed the core evidence base. Policy and institutional reports were used only when they clarified the implementation context in Iraq and Iraqi Kurdistan.

Exclusion Criteria

Studies were excluded when they discussed technology, AI, or digital learning in general terms without a clear connection to pronunciation, feedback, adaptation, learning analytics, or educational governance; when they treated ASR only as a technical engineering issue without pedagogical relevance; when they addressed speech technology outside educational or language-learning settings; or when they lacked enough conceptual or methodological detail to inform the proposed framework. The synthesis then moved from search results to

design logic. The selected studies were compared in terms of the instructional work they performed: how they explained input, feedback, adaptation, validity, fairness, and contextual feasibility. This reading exposed a central gap in current CAPT and adaptive-learning models: they often improve isolated practice or feedback delivery but rarely integrate variable input, graduated feedback, learner-specific sequencing, and governance into a single instructional pathway. The Input-Feedback-Adaptation framework was designed to address that gap and provide future empirical studies with a model they can implement, test, revise, and compare across educational settings.

THEORETICAL FOUNDATIONS

SLA Perspectives on Pronunciation Development and Instructional Targets

L2 pronunciation development involves perceptual learning, through which learners establish or reorganize phonological categories (Best & Tyler, 2007; Flege, 1995), and production learning, through which they stabilize articulatory routines and prosodic patterns that listeners can interpret reliably (Derwing & Munro, 2015; Leather & James, 1991; Saito, 2015). This evidence creates an important design constraint: a pronunciation pathway cannot rely on one narrow score, such as phoneme accuracy, because contemporary pedagogy distinguishes accentedness from comprehensibility and intelligibility and treats listener-centered outcomes as more appropriate instructional goals (Saito, 2021b). Meta-analytic evidence confirms that pronunciation instruction can improve learner outcomes, but it also shows substantial variation by instructional focus, measurement choice, and treatment design (Lee et al., 2015). An Education 4.0 pathway, therefore, needs to coordinate segmentals, suprasegmentals, timing, perception, production, and communicative transfer rather than treating pronunciation as a set of isolated drills (Amrate & Tsai, 2025).

HVPT as an Input Framework for Perception and Generalization

HVPT provides the input logic for the framework because it treats variability as a condition for learning rather than as background noise. Instead of exposing learners to a single idealized model, HVPT presents multiple talkers, tokens, phonetic contexts, and lexical environments, enabling learners to form flexible categories and generalize beyond practiced items (Mahmood, 2024b; Thomson, 2011, 2018; Wong, 2025). Recent meta-analytic work emphasizes generalization and retention as central evidence of learning, not merely immediate gains on trained items (Uchihara et al., 2025). This principle maps directly onto Education 4.0 design. A digital pathway can manage variability at scale, record which contrasts and talkers a learner has encountered, and use that record to select later input. HVPT, therefore, becomes more than a training technique; it becomes a sequencing principle for adaptive pronunciation learning.

Feedback Design: Correction and Mediation

CF supplies the second design requirement. Interactionist accounts show how feedback draws attention to form and creates opportunities for modified output, while sociocultural accounts frame feedback as a form of mediation that should match the learner's current need for support (Li et al., 2025). Meta-analytic evidence indicates that oral feedback can support L2 learning, but its effects depend on the type of feedback, the target feature, timing, learner uptake, and measurement (Lyster & Saito, 2010). In pronunciation, this means that automated feedback cannot function as simple error labeling. It must guide attention, invite repair, and adjust the degree of explicitness as learners move toward self-regulation (John et al., 2023; Liakina & Liakin, 2022; Saito, 2021a; Tejedor-García et al., 2017).

Sociocultural theory sharpens this requirement by emphasizing graduated and contingent support (Lantolf & Thorne, 2006; Vygotsky, 1978). In an AI-mediated pronunciation pathway, the system should begin with low-intrusion signals and move toward explicit correction only when learner performance, repeated attempts, or ASR confidence justify stronger mediation (Silpachai et al., 2024; Wen et al., 2023; Zhang & Zhang, 2023). This design avoids two common weaknesses in CAPT: feedback that is too vague to support repair and feedback that is too forceful when the system lacks sufficient confidence.

Adaptive Learning in Education 4.0 Systems

Adaptive learning provides the third design requirement. Education 4.0 frames digital transformation as the coordination of teaching, assessment, data, and quality assurance across connected ecosystems (Chituc, 2021). Adaptive learning operationalizes coordination by using learner data to adjust task difficulty, sequencing, feedback intensity, and long-term progression (Aleven et al., 2016). Yet adaptivity only strengthens learning when the data are valid, interpretable, and ethically governed. Learning analytics research has therefore treated privacy, consent, and fairness as central design concerns (Slade & Prinsloo, 2013). Pronunciation intensifies these concerns because ASR systems may misrecognize accented speech and produce unreliable feedback for particular speech communities (Koenecke et al., 2020). A responsible pathway must therefore model uncertainty and make adaptation explainable to teachers and learners.

Limits of Current Approaches

Current CAPT and ASR-supported pronunciation research establishes the need for a more integrated framework. CAPT can extend practice time and provide frequent feedback, but many systems still measure what technology can detect most easily, such as segmental accuracy in controlled tasks, rather than what pedagogy values most, such as intelligibility, comprehensibility, prosody, and transfer to communication (Ai, 2015; Amrate & Tsai, 2025; Cucchiarini et al., 2012; Dai & Wu, 2023; Moxon, 2024). Research also remains fragmented across single applications or short interventions, leaving three design elements weakly connected: structured input, feedback progression, and adaptive sequencing. Rogerson-Revell (2021) therefore calls for stronger alignment between pronunciation technology and pedagogy, especially through more authentic and interactive designs. ASR research adds the same point from a feedback perspective. Meta-analytic evidence shows moderate overall benefits, with stronger effects when feedback provides explicit, instructionally usable information rather than indirect signals alone (Thi-Nhu Ngo et al., 2024). At the same time, ASR accuracy varies across systems and speakers (McCrocklin et al., 2019), and documented bias in speech recognition systems raises fairness concerns (Koenecke et al., 2020). These findings do not merely summarize a gap in the literature; they underpin the logic of the proposed framework. Input must be variable and pedagogically selected, feedback must be graduated and uncertainty-aware, and adaptation must operate under ethical governance (Slade & Prinsloo, 2013).

THE PROPOSED FRAMEWORK

Overview of the Instructional Design Framework

The proposed framework conceptualizes intelligent pronunciation learning pathways (see **Figure 1** for an overview and **Figure 2** for a full visualization of the framework) as an ecosystem subsystem comprising three interacting layers. These layers emerge from the synthesis above: HVPT establishes how input should vary, CF and mediation theory establish how support should progress, and adaptive learning establishes how learner data should guide sequencing under ethical constraints.

1. Input layer (HVPT-informed exposure): curated, high-variability perceptual and production input designed to support category learning, generalization, and communicatively relevant targets (Uchihara et al., 2025).
2. Feedback layer (AI-mediated regulation): multi-channel feedback (implicit → explicit) designed as mediation and regulation, not merely error marking; includes opportunities for self-correction and reflective engagement (Thi-Nhu Ngo et al., 2024).
3. Adaptation layer (sequencing and pathway governance): learner modeling and sequencing rules that personalize exposure, tasks, and feedback intensity while managing risk (bias, privacy, over-reliance) (Aleven et al., 2016).

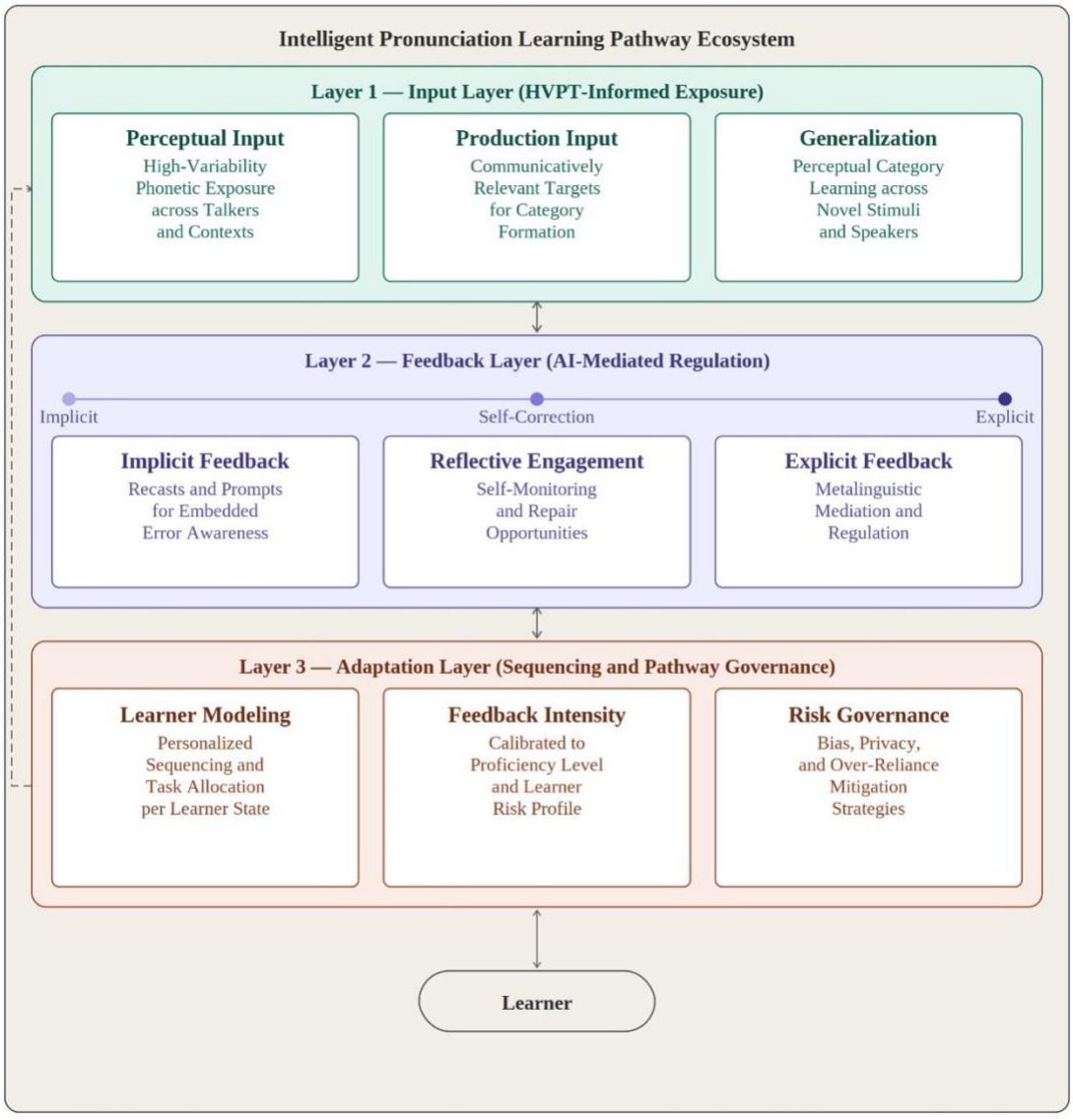

Figure 1. Instructional Design Framework for Intelligent Pronunciation Learning Pathways. Bidirectional arrows indicate recursive interaction between layers; the dashed left arrow denotes adaptive pathway feedback from Layer 3 to Layer 1. (Source: Author's original conceptual framework; AI-assisted visualization using Claude 4.6.)

Crucially, these layers must be designed as a dynamic system rather than a pipeline. HVPT-style input alters what learners can perceive; perception influences production; production affects how ASR interprets learner speech; ASR interpretation shapes the feedback learners receive; and feedback changes learner strategies, engagement, and later performance data. A pathway, therefore, requires explicit design of interactions and fail-safes (Aleven et al., 2016). The following sections present the core components of the proposed framework.

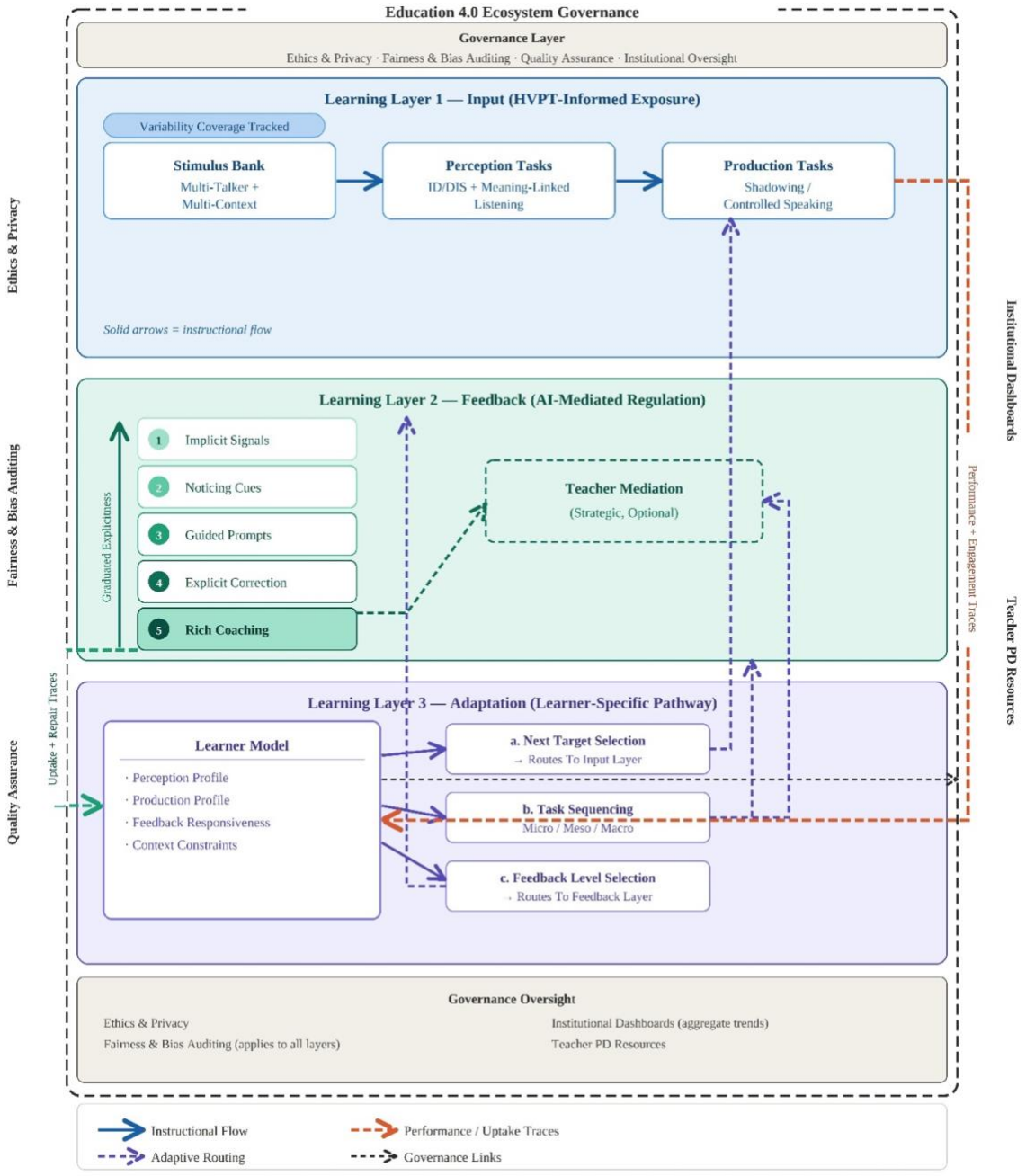

Figure 2. Intelligent Pronunciation Learning Pathways in Education 4.0: Input–Feedback–Adaptation Layers and Ecosystem Governance. (Source: Author's original conceptual framework; AI-assisted visualization using Claude 4.6)

Novelty Relative to CAPT and Adaptive Learning Models

The framework differs from existing CAPT and adaptive learning models in three ways. First, it treats pronunciation learning as a pathway rather than a tool, so the design unit is the relation among input, feedback, adaptation, teacher mediation, and governance. Second, it makes pronunciation-specific data central to adaptation: perception confusions, production patterns, ASR confidence, repair after feedback, and generalization across tasks all inform sequencing decisions. Third, it places fairness and privacy inside the instructional architecture rather than treating them as external ethical reminders. In this sense, the framework is not another ASR application; it is a design grammar for coordinating pronunciation content, automated feedback, teacher judgment, and contextual constraints (Aleven et al., 2016; Koenecke et al., 2020; Slade & Prinsloo, 2013; Uchihara et al., 2025).

A Three-Step Process for Intelligent Pronunciation Learning

The proposed framework can be understood as a continuous learning cycle rather than a complex technical system. Instead of presenting multiple variables at once, the process unfolds in three clear stages: Input → Feedback → Adaptation. Each stage builds on the previous one and feeds into the next.

Step 1: Input: Building Perception Through Structured Variability

The learning process begins with well-designed input. Instead of presenting one "correct" pronunciation, the system exposes learners to multiple versions of the same sound across different speakers and contexts. This approach, based on High Variability Phonetic Training (HVPT), helps learners build more flexible and accurate sound categories (Uchihara et al., 2025). Learning then moves through a clear sequence:

- Learners first identify and discriminate sounds (e.g., distinguishing /iː/ and /ɪ/)
- They then move to guided production tasks, such as repetition or shadowing
- Finally, they engage in meaning-focused tasks, where pronunciation affects understanding

Importantly, variability is not random. The system tracks what the learner has already encountered and gradually introduces new combinations of speakers and contexts to ensure balanced exposure. The outcome of this stage is twofold:

- Learners develop a clearer perception of sounds
- The system collects data about which sounds are difficult and underdeveloped

Step 2: Feedback: Guiding Attention and Self-Correction

Once learners produce speech, the system provides feedback. However, rather than immediately correcting every error, feedback is delivered gradually based on the learner's needs. This process follows a five-level progression:

1. Implicit signals: The system simply indicates a mismatch (e.g., incorrect transcription) without explanation
2. Noticing support: It highlights the problematic part or contrasts it with a correct model
3. Guided prompts: The learner receives hints such as "Try a longer vowel."
4. Explicit correction: The correct form is provided with minimal explanation
5. Detailed explanation: Full articulatory guidance or structured practice is given

The goal of this progression is to support self-correction first and increase explicit feedback only when needed. This approach reflects research on CF and the idea of learning through guided support (Lyster & Saito, 2010; Zhang & Zhang, 2023). At the same time, feedback must be reliable. When the system is uncertain, such as in the case of ASR errors, it should avoid making strong corrections and instead provide guidance. This helps prevent misleading learners, especially when speech recognition may be biased or inaccurate (Koenecke et al., 2020).

Step 3: Adaptation: Personalizing the Learning Pathway

The final stage uses learner data to adjust future learning. This is where the system becomes "*intelligent*." The system continuously analyzes:

- Which sounds the learner struggles with
- How they respond to feedback
- Whether they improve after correction
- How consistently they perform across tasks

Adaptation Layer: Learner-Specific Pathways and Dynamic Interactions

Adaptive learning responds to learner data at different time scales, from immediate adjustments to longer-term sequencing decisions (Aleven et al., 2016). In pronunciation learning, this is essential because learners vary in the sounds they find difficult, their responses to feedback, and the speed at which they improve. Research on CAPT also shows that fixed sequences and repetitive drills are not effective. This makes adaptivity necessary for efficient and fair learning at the system level (Amrate & Tsai, 2025).

Learner Model: Key Tracking Components

- Perception profile: confusion matrices for contrasts; generalization performance across talker/context variability (Uchihara et al., 2025).
- Production profile: target-specific error rates (e.g., segmental substitutions, stress misplacement), with uncertainty tagging when ASR confidence is low (Thi-Nhu Ngo et al., 2024).
- Feedback responsiveness: whether prompts lead to improvement; time to repair; persistence patterns (Gooch et al., 2016).
- Engagement and persistence proxies: completion rates, repeated attempts, dropout from tasks, treated cautiously and ethically (Slade & Prinsloo, 2013).
- Context constraints: Learners also face a range of contextual limitations, including the type of device they are using, whether they have a fast enough internet connection, and what restrictions their institution has put in place (Chituc, 2021).

Adaptive Sequencing: Control Mechanisms

1) Micro-adaptation (within session): adjusts feedback level, repetition, and input based on immediate learner performance. If errors continue or confidence is low, support becomes more explicit (Zhang & Zhang, 2023).

2) Meso-adaptation (across sessions): organizes practice over time by spacing tasks, varying speakers and contexts, and selecting activities based on learner needs (perception vs. production). This supports steady learning progress (Aleven et al., 2016).

3) Macro-adaptation (long-term pathway): shifts learning goals over time and aligns with classroom instruction. It also tracks overall progress and identifies where support or resources are needed (Chituc, 2021).

Dynamic Interaction: Core Design Principle

- HVPT input shapes how learners hear sounds, which affects how they speak and how ASR evaluates their speech. This means feedback accuracy depends partly on how the input is designed (Uchihara et al., 2025).
- Feedback influences how learners approach tasks, which changes the data used for adaptation. This means adaptation depends partly on feedback design (McCrocklin et al., 2019).
- Adaptation controls what input learners receive, which affects how well they generalize new sounds. This means HVPT effectiveness depends partly on adaptation decisions.

SYNTHESIS FOR EDUCATION 4.0 IMPLEMENTATION

Instructional Design Principles

The framework translates the synthesis into six instructional design principles for Education 4.0 pronunciation learning. First, pronunciation should be integrated into the wider learning system, rather than a peripheral application detached from teaching, assessment, and teacher development (Chituc, 2021). Second, input should be deliberately variable, with the system tracking talkers, tokens, contexts, and target contrasts rather than presenting isolated models (Uchihara et al., 2025). Third, feedback should progress from noticing support to explicit correction only when learner performance warrants such a shift (Thi-Nhu Ngo et al., 2024). Fourth, automated correction should remain uncertainty-aware; when ASR confidence is low, the system should guide learners rather than make strong claims of error (Koenecke et al., 2020). Fifth, adaptation should remain explainable and fair, allowing learners and teachers to understand why tasks appear and how data shape later practice (Slade & Prinsloo, 2013). Finally, teachers should remain central because they interpret data, connect AI feedback to classroom goals, and prevent automated feedback from replacing pedagogical judgment (Chituc, 2021).

Contextualizing the Framework in Iraqi Kurdistan

Iraqi Kurdistan is not treated here as a background example but as a test of the framework's ecological validity. In Iraq, national education policy emphasizes quality improvement, system strengthening, and preparation for modern labor-market demands (Ministry of Education & World, 2024). Yet implementation remains constrained by uneven infrastructure, limited access to reliable technology, and gaps in digital readiness. Research on e-learning in Iraqi higher education has reported persistent barriers related to infrastructure, teacher preparation, and institutional implementation (Al-Azawei et al., 2016). These conditions challenge any Education 4.0 model that assumes stable connectivity, abundant devices, or high teacher confidence with digital tools.

The same constraints become especially consequential for pronunciation. Kurdish EFL teachers may have strong pedagogical and content knowledge, yet many report limited confidence in technology-supported teaching (Ali & Mohammadzadeh, 2022; Mahmood, 2025). Pronunciation teaching also suffers from limited classroom time, few locally available tools, and insufficient teacher training (Mahmood, 2024a). Broader evidence on English-language teaching in Iraq similarly shows that sustainable improvement requires teacher development, institutional support, and system-level reform rather than updated materials alone (Borg & Capstick, 2024). The Input-Feedback-Adaptation framework, therefore, needs contextual translation: input

must be downloadable and locally usable, feedback must remain teacher-mediated when ASR reliability is uncertain, and adaptation must work with sparse data and uneven access.

Implementation Conditions for Iraqi Kurdistan

These conditions lead to five implementation requirements. First, the input layer should support low-bandwidth and offline access, allowing learners to download HVPT materials, practice locally, and upload performance data when a connection becomes available (Chituc, 2021). Second, the feedback layer should assist rather than replace teachers. In low-resource classrooms, AI feedback can triage common pronunciation problems, while teachers concentrate on learners who need explanation, modeling, or reassurance. Third, teacher development should focus on pronunciation-specific digital pedagogy: setting intelligibility-oriented goals, interpreting ASR feedback, selecting HVPT materials, and linking practice to communicative tasks (Ali & Mohammadzadeh, 2022). Fourth, the adaptation layer should use limited data responsibly, collect only what the pathway needs, process speech locally where possible, and give learners clear control over data use (Slade & Prinsloo, 2013). Fifth, implementation should proceed through small pilots before institutional scaling, because infrastructure and policy support remain uneven (United Nations Development, 2024).

DISCUSSION

The proposed framework aligns pronunciation pedagogy with Education 4.0 by shifting the unit of design from individual tools to learning pathways. Its contribution does not lie in inventing HVPT, CF, ASR, or adaptive learning; these elements already have substantial research bases. Its contribution lies in specifying how these elements should interact: input provides structured variability; feedback mediates attention and repair; adaptation sequences later practice; and governance protects learners when data and ASR outputs remain imperfect. This synthesis responds to evidence that pronunciation instruction can work but varies by design (Lee et al., 2015), that ASR-supported pronunciation learning benefits from usable feedback (Thi-Nhu Ngo et al., 2024), and that CAPT still struggles to connect technological affordances with communicative pedagogy (Amrate & Tsai, 2025). Four tensions define the framework's usefulness. The first concerns automation and validity. Speech technologies can provide frequent feedback, but they can also deliver inaccurate guidance when ASR confidence is low or when accent variation falls outside the system's training distribution. The framework addresses this problem through uncertainty-aware feedback: low-confidence outputs trigger prompts, comparisons, or teacher review rather than categorical error claims (Koenecke et al., 2020). The second tension concerns learner agency. Adaptive systems can efficiently sequence learning, but opaque recommendations can reduce learner autonomy. The framework, therefore, requires explainable adaptation, teacher override, and transparent data governance (Slade & Prinsloo, 2013). The third tension concerns assessment. CAPT systems often privilege segmental accuracy because it can be measured automatically, whereas pronunciation pedagogy increasingly values intelligibility and comprehensibility. A multi-layer assessment design can use automated measures for frequent practice while relying on teacher, peer, or listener-based checks to maintain communicative validity (Amrate & Tsai, 2025). The fourth tension concerns context. Iraqi Kurdistan shows why Education 4.0 frameworks cannot assume ideal infrastructure. In such settings, success should mean pedagogically guided integration: AI extends opportunities for practice, but teachers still set goals, interpret feedback, and protect learners from misleading automated judgments (Ali & Mohammadzadeh, 2022).

Limitations of the Conceptual Study

The framework has the limitations of a conceptual study. It does not report learner performance, teacher uptake, implementation cost, or system accuracy data. Its literature selection was purposive and theory-driven rather than systematic, so it should not be read as a complete evidence map of CAPT, ASR, HVPT, or adaptive

learning research. Iraqi Kurdistan functions as an illustrative context, not as an empirical case study; therefore, the framework cannot claim contextual validity until researchers test it with local teachers, learners, devices, institutions, and data-protection practices. Future studies should examine whether the pathway improves perception, production, intelligibility, learner engagement, teacher decision-making, and ASR fairness across proficiency levels and accent backgrounds.

CONCLUSION

This study proposed an instructional design framework for intelligent pronunciation learning pathways within Education 4.0 ecosystems. The framework integrates HVPT-informed input, AI-mediated feedback as graduated support, and adaptive sequencing based on learner modeling, while placing privacy, fairness, uncertainty, and teacher mediation inside the design rather than after implementation. The framework responds to research showing that pronunciation instruction can improve learner outcomes, that ASR-supported pronunciation learning depends on feedback design, and that CAPT research still needs stronger pedagogical integration. The discussion of Iraqi Kurdistan shows that intelligent pronunciation systems must operate under real institutional constraints, including uneven infrastructure, limited pronunciation resources, and teacher development needs. The framework should now be tested through design-based studies, classroom pilots, cross-institutional comparisons, and fairness audits of ASR-supported feedback. Pronunciation should not remain an optional supplement in digital language education; it should become a designed, measurable, and ethically governed component of communicative learning.

Funding: No funding was received to assist with the preparation of this manuscript.

Declaration of interest: The authors declare no competing or conflicting interests.

Ethical statement: Research did not involve human participants.

AI statement: Grammarly was used exclusively for language editing, including improvements to grammar, spelling, punctuation, sentence structure, and readability. Claude 4.6 was used solely to generate visual representations of the author's original conceptual framework according to the author's specifications. All AI-assisted outputs were reviewed and verified by the author, who retains full responsibility for the accuracy, originality, and integrity of the manuscript.

Data sharing statement: No data was collected for this study.

REFERENCES

Ai, R. L. (2015). Automatic pronunciation error detection and feedback generation for CALL applications. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies, LCT 2015* (pp. 175-186), https://doi.org/10.1007/978-3-319-20609-7_17

Al-Azawei, A., Parslow, P., & Lundqvist, K. (2016). Barriers and opportunities of e-learning implementation in Iraq: A case of public universities. *The International Review of Research in Open and Distributed Learning*, *17*(5), 126-146. https://doi.org/10.19173/irrodl.v17i5.2501

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In V. Aleven, E. A. McLaughlin, R. A. Glenn, & K. R. Koedinger (Eds.), *Handbook of research on learning and instruction* (pp. 477-518). Routledge. https://doi.org/10.4324/9781003484776-26

Ali, S. S., & Mohammadzadeh, B. (2022). Iraqi Kurdish EFL teachers' beliefs about technological pedagogical and content knowledge: The role of teacher experience and education. *Frontiers in Psychology*, *13, Article 969195*. https://doi.org/10.3389/fpsyg.2022.969195

Almusharraf, A., Mahdi, H. S., Al-Nofaie, H., & Aljasser, A. (2024). The effect of settings, educational level and tools on computer-assisted pronunciation training: A meta-analysis. *Journal of Computer Assisted Learning*, *40*(4), 1605-1615. https://doi.org/10.1111/jcal.12974

Amrate, M., & Tsai, P.-h. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, *37*(1), 22-42. https://doi.org/10.1017/S0958344024000181

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Borg, S., & Capstick, T. (2024). *Iraq – Understanding English language teaching and learning*. British Council. https://doi.org/10.57884/P0XB-4S40

Chituc, C.-M. (2021). A framework for Education 4.0 in digital education ecosystems. In L. M. Camarinha-Matos, X. Boucher, & H. Afsarmanesh (Eds.), *Working conference on virtual enterprises, 2021* (pp. 702-709), Cham. https://ifip.hal.science/emse-03350033/

Crowther, D., & Loewen, S. (2025). Instructed second language acquisition and second language pronunciation. *Language Learning*, *75*, 30-54. https://doi.org/10.1111/lang.12716

Cucchiarini, C., Nejjari, W., & Strik, H. (2012). My Pronunciation Coach: Improving English pronunciation with an automatic coach that listens. *Language Learning in Higher Education*, *1*(2), 365-376. https://doi.org/10.1515/cercles-2011-0024

Dai, Y. J., & Wu, Z. W. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, *36*(5-6), 861-884. https://doi.org/10.1080/09588221.2021.1952272

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://doi.org/10.1075/lllt.42

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). York Press. https://www.researchgate.net/publication/333815781_Second_language_speech_learning_Theory_findings_and_problems

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System*, *60*, 117-127. https://doi.org/10.1016/j.system.2016.06.007

Hirschi, K., Kang, O., Yang, M., Hansen, J. H. L., & Beloin, K. (2025). Artificial intelligence-generated feedback for second language intelligibility: An exploratory intervention study on effects and perceptions. *Language Learning*, *75*, 204-241. https://doi.org/10.1111/lang.12719

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1-2), 18-26. https://doi.org/10.1007/s13162-020-00161-0

John, P., Cardoso, W., & Johnson, C. (2023). Automatic speech recognition as a source of corrective feedback on L2 pronunciation. In P. John, W. Cardoso, C. Johnson (Eds.), *Frontiers in technology-mediated language learning* (pp. 41-56). Routledge. https://doi.org/10.4324/9781003395218-4

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684-7689. https://doi.org/doi:10.1073/pnas.1915768117

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press. https://books.google.com.au/books?id=HlVpAAAAMAAJ&hl=et&source=gbs_book_other_versions_r&cad=3&redir_esc=y

Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition*, *13*(3), 305-341. https://doi.org/10.1017/S0272263100010019

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366. https://doi.org/10.1093/applin/amu040

Li, L., Zhang, X., Zou, B., & Yang, Q. (2025). AI partner or peer partner? Exploring AI-mediated interaction in EFL pronunciation from a socio-cultural perspective. *Learning, Culture and Social Interaction*, *55*, Article 100958. https://doi.org/10.1016/j.lcsi.2025.100958

Liakina, N., & Liakin, D. (2022). Speech technologies and pronunciation training: What is the potential for efficient corrective feedback? In U. K. Alves and J. I. A. de Albuquerque (Eds.), *Second language pronunciation: Different approaches to teaching and training* (pp. 287-314). De Gruyter. https://doi.org/10.1515/9783110736120-011

Liu, Y., Ab Rahman, F. B., & Zain, F. B. M. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, *12*(1), Article 2466288. https://doi.org/10.1080/2331186X.2025.2466288

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*(2), 265-302. https://doi.org/10.1017/S0272263109990520

Mahmood, R. Q. (2024a). ESL learners' perceptions of using high-variability phonetic training (HVPT) through YouGlish to improve pronunciation skills. *Australian Journal of Applied Linguistics*, *7*(3), Article 1818. https://doi.org/10.29140/ajal.v7n3.1818

Mahmood, R. Q. (2024b). Teaching pronunciation in EFL classes: An investigative study among Kurdish EFL teachers. *MEXTESOL Journal*, *48*(4), 1-12. https://doi.org/10.61871/mj.v48n4-3

Mahmood, R. Q. (2025). Exploring pedagogical effectiveness in teacher education: Insights and challenges from Kurdish EFL university lecturers. *Journal of Interdisciplinary Studies in Education*, *14*(4), 1-26. https://ojed.org/jise/article/view/7706

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 191-200), Ames. https://www.researchgate.net/publication/334001187_ASR_dictation_program_accuracy_Have_current_programs_improved

Ministry of Education, I., & World, B. (2024). *National education strategy for Iraq 2022-2031*. https://documents1.worldbank.org/curated/en/099091824091030738/pdf/P171165-cfac2b5e-6448-4093-a5a5-2591a8024209.pdf

Moxon, S. (2024). All-talk: Enhancing EFL pronunciation with Microsoft Azure Speech Services. *ABAC Journal*, *44*(4), 139-161. https://doi.org/10.59865/abacj.2024.58

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, *52*(1), 189-205. https://doi.org/10.1177/0033688220977406

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563-595. https://doi.org/10.1111/lang.12120

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

Saito, K. (2021b). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866-900. https://doi.org/10.1002/tesq.3027

Silpachai, A., Neiriz, R., Novotny, M., Gutierrez-Osuna, R., Levis, J. M., & Chukharev, E. (2024). Corrective feedback accuracy and pronunciation improvement: Feedback that is 'good enough'. *Language Learning & Technology*, *28*(1), 1-16. https://www.lltjournal.org/item/10125-73582/

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

Souza, A. S. C. d., & Debs, L. (2024). Concepts, innovative technologies, learning approaches and trend topics in education 4.0: A scoping literature review. *Social Sciences & Humanities Open*, *9*, Article 100902. https://doi.org/10.1016/j.ssaho.2024.100902

Tejedor-García, C., Escudero-Mancebo, D., González-Ferreras, C., Cámara-Arenas, E., & Cardeñoso-Payo, V. (2017). Evaluating the efficiency of synthetic voice for providing corrective feedback in a pronunciation training tool based on minimal pairs. *7th ISCA Workshop on Speech and Language Technology in Education, SLaTE 2017*, 25-29, https://doi.org/10.21437/SLaTE.2017-5

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4-21. https://doi.org/10.1017/S0958344023000113

Thomson, R. I. (2011). Computer-Assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744-765. https://doi.org/10.11139/cj.28.3.744-765

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208-231. https://doi.org/10.1075/jslp.17038.tho

Uchihara, T., Karas, M., & Thomson, R. I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*, *47*(3), 794-827. https://doi.org/10.1017/S0272263125100879

United Nations Development, P. (2024). *Country programme document for Iraq (2025-2029)* (DP/DCP/IRQ/4). https://www.undp.org/sites/g/files/zskgke326/files/2024-12/iraq-cpd.pdf

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://www.hup.harvard.edu/books/9780674576292

Wen, Y. R., Li, J., Xu, H. K., & Hu, H. W. (2023). Restructuring multimodal corrective feedback through augmented reality (AR)-enabled videoconferencing in L2 pronunciation teaching. *Language Learning & Technology*, *27*(3), 83-107. https://doi.org/10.64152/10125/73533

Wong, J. W. S. (2025). What do the stimuli in high variability phonetic training tell us about second language perception? Current findings, implications, and the way forward. In M. Reed, J. M. Levis (Eds.), *The handbook of second language listening* (pp. 155-168). Wiley. https://doi.org/10.1002/9781394312375.ch12

Zhang, X., & Zhang, R. (2023). Feedback, response, and learner development: A sociocultural approach to corrective feedback in second language writing. *Sage Open*, *13*(1), Article 21582440231157680. https://doi.org/10.1177/21582440231157680

APPENDIX A

Studies Retrieved and Read for the Conceptual Theory Synthesis

Search scope. The same keyword logic was used across Web of Science and Scopus, with database-specific field codes only. Web of Science was searched in the Topic field for 2010-2026 and returned 12 records. Scopus was searched in TITLE-ABS-KEY and limited to the Social Sciences and Arts and Humanities subject areas, returning 11 records. All retrieved records were read for relevance to *the Input-Feedback-Adaptation framework*.

Master keyword string. ("second language pronunciation" OR "L2 pronunciation" OR "EFL pronunciation" OR "ESL pronunciation") AND ("computer-assisted pronunciation training" OR CAPT OR "automatic speech recognition" OR ASR OR "AI-mediated feedback" OR "AI feedback") AND ("high variability phonetic training" OR HVPT OR "corrective feedback" OR "adaptive learning" OR "learning analytics" OR "personalized learning" OR "Education 4.0").

Screening note. The two database exports contained 23 records in total. After removing overlapping records across databases, 17 unique studies/records remained. These records were used as the search-based appendix for the conceptual synthesis; additional theoretical and contextual sources cited in the manuscript were used where they clarified the framework or the Iraqi Kurdistan implementation context. One study was excluded as it was not written in English from the Web of Science.

Table A1. Search yield

Database	Field/search format	Limits	Records retrieved	Records read
Web of Science	Topic field (TS)	2010-2026	12	11
Scopus	TITLE-ABS-KEY	Social Sciences; Arts and Humanities	11	11
Total database records	-	-	23	22
Unique records after overlap check	-	-	18	17

Table A2. Unique records extracted from Web of Science and Scopus

No.	Database source	Year	Authors	Title	Publication/source and type	DOI
1	Web of Science	2025	Crowther, D.; Loewen, S.	Instructed Second Language Acquisition and Second Language Pronunciation	Language Learning (Review)	10.1111/lang.12716
2	Web of Science	2025	Hirschi, K.; Kang, O.; Yang, M.; Hansen, J. H. L.; Beloin, K.	Artificial Intelligence-Generated Feedback for Second Language Intelligibility: An Exploratory Intervention Study on Effects and Perceptions	Language Learning (Article)	10.1111/lang.12719
3	Web of Science; Scopus	2025	John, P.; Johnson, C.; Cardoso, W.	Exploring automatic speech recognition for corrective and confirmative pronunciation feedback	Journal of Second Language Pronunciation (Article)	10.1075/jslp.24035.joh

No.	Database source	Year	Authors	Title	Publication/source and type	DOI
4	Web of Science	2025	Liu, Y.; Ab Rahman, F. B.; Zain, F. B. M.	A systematic literature review of research on automatic speech recognition in EFL pronunciation	Cogent Education (Review)	10.1080/2331186x.2025.2466288
5	Scopus	2025	WONG, J.W.S.	What Do the Stimuli in High Variability Phonetic Training Tell Us About Second Language Perception? Current Findings, Implications, and the Way Forward	The Handbook of Second Language Listening (Book chapter)	10.1002/9781394312375.ch12
6	Web of Science	2024	Almusharraf, A.; Mahdi, H. S.; Al-Nofaie, H.; Aljasser, A.	The effect of settings, educational level and tools on computer-assisted pronunciation training: A meta-analysis	Journal of Computer Assisted Learning (Article)	10.1111/jcal.12974
7	Scopus	2024	Moxon, S.	All-talk: enhancing efl pronunciation with microsoft azure speech services	ABAC Journal (Article)	10.59865/abacj.2024.58
8	Web of Science; Scopus	2024	Ngo, T. T. N.; Chen, H. H. J.; Lai, K. K. W.	The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis	ReCALL (Article)	10.1017/s0958344023000113
9	Web of Science; Scopus	2024	Silpachai, A.; Neiriz, R.; Novotny, M.; Gutierrez-Osuna, R.; Levis, J. M.; Chukharev, E.	Corrective feedback accuracy and pronunciation improvement: Feedback that is 'good enough'	Language Learning & Technology (Article)	10.64152/10125/73582
10	Web of Science	2023	Dai, Y. J.; Wu, Z. W.	Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study	Computer Assisted Language Learning (Article)	10.1080/09588221.2021.1952272
11	Scopus	2023	John, P.; Cardoso, W.; Johnson, C.	Automatic speech recognition as a source of corrective feedback on L2 pronunciation	Frontiers in Technology-Mediated Language Learning (Book chapter)	10.4324/9781003395218-4
12	Web of Science; Scopus	2023	Wen, Y. R.; Li, J.; Xu, H. K.; Hu, H. W.	Restructuring multimodal corrective feedback through Augmented Reality (AR)-enabled videoconferencing in L2 pronunciation teaching	Language Learning & Technology (Article)	10.64152/10125/73533

No.	Database source	Year	Authors	Title	Publication/source and type	DOI
13	Scopus	2022	Liakina, N.; Liakin, D.	Speech technologies and pronunciation training: What is the potential for efficient corrective feedback?	Second Language Pronunciation: Different Approaches to Teaching and Training (Article)	10.1515/9783110736120-011
14	Scopus	2017	Tejedor-García, C.; Escudero-Mancebo, D.; González-Ferreras, C.; Cámara-Arenas, E.; Cardeñoso-Payo, V.	Evaluating the Efficiency of Synthetic Voice for Providing Corrective Feedback in a Pronunciation Training Tool Based on Minimal Pairs	7th ISCA Workshop on Speech and Language Technology in Education, SLaTE 2017 (Conference paper)	10.21437/slate.2017-5
15	Scopus	2016	Detey, S.; Fontan, L.; Pellegrini, T.	Pronunciation processing in a foreign language: Didactic approaches, automatic methods and challenges for learning	Revue Traitement Automatique des Langues (not English)	https://www.atala.org/content/tal_57_3_1
16	Web of Science	2015	Ai, R. L.	Automatic Pronunciation Error Detection and Feedback Generation for CALL Applications	Learning and Collaboration Technologies, LCT 2015 (Proceedings Paper)	10.1007/978-3-319-20609-7_17
17	Web of Science	2012	Cucchiarini, C.; Nejjari, W.; Strik, H.	My Pronunciation Coach: Improving English pronunciation with an automatic coach that listens	Language Learning in Higher Education (Article)	10.1515/cercles-2011-0024
18	Web of Science; Scopus	2011	Thomson, R.I.	Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation	CALICO Journal (Article)	10.11139/cj.28.3.744-765

Note. WoS = Web of Science. Duplicate records across WoS and Scopus were merged in Table A2 and marked with both database sources. The table reports the records extracted from the database exports supplied for the conceptual theory synthesis.