RESEARCH IN ENGLISH LANGUAGE EDUCATION JOURNAL
E-ISSN 3115-7599

Dissecting the Pedagogical Gap: An Empirical Metacognitive Study of Vision 3's Grammar and Iran's National English Exam (2022–2025)

Ehsan Khaef [1] & Mohammad Hasan Razmi [2]

[1] Fars Education Department, Ministry of Education, Fasa, Iran.

ehsankhaef2006@gmail.com

[2] *Corresponding Author:* Department of English Language Teaching, Farhangian University, P. O. Box 14665-889, Tehran, Iran.

mrazmi@cfu.ac.ir

ABSTRACT

This study critically examines the pedagogical disparity between Vision 3's grammar instruction and the cognitive and metacognitive demands of Iran's national English examinations (2022–2025), assessing its efficacy in preparing students for high-stakes assessments pivotal to Konkoor outcomes. Hypothesizing Vision 3 prioritizes lower-order cognitive tasks and faces systemic barriers, the study uses a mixed-methods design. Quantitative content analysis, guided by Revised Bloom's Taxonomy, evaluated 620 grammar items across six topics (passive voice, tag questions, relative clauses, conditional sentences Type II, passive voice with modals, past perfect tense), followed by interviews with 20 students and 10 teachers. Findings reveal Vision 3's focus on lower-order skills (68% Understanding/Applying, 15% metacognitive prompts) and mechanical tasks (72%), contrasting with exams' emphasis on higher-order skills (65% Analyzing/Evaluating, 60% metacognitive prompts) and analytical tasks (68%). Stakeholders reported Vision 3 as inadequate, necessitating metacognitive strategies (e.g., self-testing, online resources). Systemic constraints—rushed syllabi, holiday disruptions, and progression-driven policies—exacerbate inequities for state school students. These results challenge Vision 3's alignment with exam demands, violating curriculum coherence and scaffolding principles. The study advocates curriculum reforms integrating analytical tasks and metacognitive scaffolding to enhance exam preparedness and educational equity in Iran's EFL context.

ARTICLE INFO:
Received: 2026-02-05
Revised: 2026-05-24
Accepted: 2026-06-24
Published online:2026-10-12

Keywords:
Pedagogical disparity, curriculum alignment, cognitive taxonomy, metacognitive scaffolding, educational equity

Article type: Research Article | **Publisher**: Farhangian University https://elt.cfu.ac.ir
©2026/The author(s) retain the copyright and full publishing rights

Citation: Khaef, E. & Razmi, M. H. (2026). Dissecting the pedagogical gap: An empirical metacognitive study of Vision 3's grammar and Iran's National English Exam (2022–2025). *Research in English Language Education Journal*, *5*(1), 39-67. https://doi.org/10.48310/relej.2026.22318.1184

1. Introduction

In Iran's centralized educational system, the *Vision 3* textbook, the culminating volume of the *Vision* series introduced nearly a decade ago, serves as the primary resource for 12th-grade English language instruction, preparing students for the high-stakes national English examination, a critical component influencing their final results in the *Konkoor* university entrance examination. Designed with a communicative-based approach, *Vision 3* prioritizes interactive language skills, with grammar relegated to a secondary role within its pedagogical framework. The presentation of grammar content is structured through two key sections: the "Grammar" section and the "See Also" part. The "Grammar" section employs an inductive-deductive hybrid approach, presenting extracts where target grammar points are boldfaced to elicit student inference, followed by explicit charts delineating grammatical rules and concluding with simple, straightforward drills. Conversely, the "See Also" part, limited to a single page, introduces additional grammar points through boldfaced sample sentences without explanatory commentary, offering minimal scaffolding. Both sections direct students to supplementary exercises in the *Vision 3* workbook, which, while occasionally challenging, fall short of the cognitive and metacognitive rigor demanded by the grammar questions in the national English exam.

This structure reveals a significant pedagogical gap, characterized by a profound mismatch between the grammar presentation in *Vision 3* and the rigorous demands of the national English exam, which significantly influences students' *Konkoor* outcomes. The exam's grammar questions require complex analytical skills and strategic reasoning, far exceeding the simplicity of *Vision 3*'s drills and workbook exercises, which primarily target basic comprehension and application. This mismatching difficulty leaves students, particularly those without prior English exposure, ill-equipped to meet the exam's expectations. Compounding this issue is the severe time constraint imposed by the *Vision 3* syllabus, which spans over 100 pages and must be covered in less than six months. Frequent disruptions from unexpected holidays further compress instructional time, rendering comprehensive grammar teaching nearly impossible. Reports indicate that even highly studious students, who excel in other subjects, often achieve low grades in English, especially those relying solely on the school-based curriculum without access to external language institutes (Sadeghi & Richards, 2016). This gap disproportionately affects students without supplementary resources, highlighting an educational inequity rooted in the curriculum's inability to align with the exam's rigor.

Prior studies on Iranian EFL textbooks underscore this misalignment. Jamshidian et al. (2024) found that the *Vision* series inadequately presents grammatical knowledge, limiting students' ability to handle complex linguistic tasks, a finding directly relevant to *Vision 3*'s grammar presentation. Similarly, Azizifar et al. (2010) noted that Iranian high school ELT textbooks historically prioritize reading comprehension over comprehensive grammar instruction, contributing to a persistent gap with national assessments. Atai and Mazlum (2013) argue that Iran's English curriculum struggles to balance communicative goals with the grammar-heavy demands of standardized tests, a tension evident in *Vision 3*'s design. Additionally, Razmjoo and Mirzaee (2016) highlight the critical role of metacognitive strategies in EFL instruction, noting that Iranian students often lack opportunities to develop self-regulated learning skills

necessary for navigating complex exam questions. Furthermore, students' foundational weaknesses in English stem from multiple factors, most notably a lack of educational justice and the substandard quality of education, particularly in state schools. Recent Ministry of Education policies exacerbate this issue by permitting students to advance to higher grades despite failing subject scores, resulting in knowledge deficits that fail to meet *Vision 3*'s course requirements (Tohidian et al., 2023; Mohammadi, 2022). For instance, Tohidian et al. (2023) highlight how systemic inequities, including underfunded state schools and inadequate teacher training, hinder effective English instruction, while Mohammadi (2022) critiques policies that prioritize progression over mastery, leaving students unprepared for high-stakes assessments like the national English exam. Despite these insights, to the best of the researcher's knowledge, no study has specifically investigated the pedagogical gap between *Vision 3*'s grammar presentation and the national English exam's rigor through an empirical and metacognitive lens, making this study a novel contribution to the field.

This study employs a mixed-methods explanatory sequential design, integrating quantitative content analysis of *Vision 3* grammar presentation and 2022–2025 exam questions with qualitative interviews of 12th-grade students and teachers to examine perceptions and strategies. The study is guided by three research questions: (1) To what extent does the presentation of grammar content in *Vision 3*—including explanations, examples, and exercises—align with the cognitive and metacognitive demands of grammar questions in Iran's national English exams (2022–2025), and how does this alignment impact student preparedness? (2) How do 12th-grade students and their teachers perceive and navigate the pedagogical gap between *Vision 3* grammar content and the rigor of Iran's national English exam questions (2022–2025), particularly through the metacognitive strategies they employ during preparation and test-taking? (3) What curriculum-related factors in Iran's educational system contribute to the misalignment between *Vision 3* grammar content and the rigor of national English exam questions (2022–2025), and how do these factors influence teachers' metacognitive instructional strategies? By addressing these questions, this study aims to elucidate the pedagogical and systemic factors undermining student performance and propose strategies for aligning curriculum and assessment to enhance educational equity in Iran's EFL context.

2. Review of Literature

This literature review synthesizes theoretical and empirical studies relevant to the pedagogical gap between grammar instruction in Iran's *Vision 3* textbook and the cognitive and metacognitive demands of the national English examination, a critical component influencing *Konkoor* outcomes. The review is structured into two sections: theoretical studies, which provide frameworks for analyzing cognitive demands, metacognitive strategies, instructional scaffolding, and curriculum alignment; and empirical studies, which evaluate *Vision 3*'s grammar presentation, students' exam challenges, and systemic curriculum issues in Iran's EFL context. By connecting individual learning processes to broader educational systems, the review establishes a foundation for understanding how *Vision 3*'s grammar teaching fails to prepare students for national exams, highlighting the need for an integrated empirical and metacognitive approach to address this gap.

2-1. Theoretical Studies

Theoretical studies offer frameworks for dissecting the pedagogical gap in *Vision 3* grammar teaching, emphasizing cognitive development, metacognitive self-regulation, instructional scaffolding, and curriculum coherence. These theories logically connect individual learner challenges to systemic issues in Iran's EFL education, providing tools to analyze alignment with national exam rigor.

Anderson and Krathwohl's (2001) Revised Bloom's Taxonomy is a pivotal framework for assessing cognitive demands in educational materials. It categorizes learning into six cognitive process levels—Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating—across four knowledge types (factual, conceptual, procedural, metacognitive). In the context of *Vision 3*, this taxonomy is crucial for evaluating whether grammar exercises foster higher-order thinking skills (HOTS) like analyzing sentence structures, which national exams demand, or remain at lower levels like remembering rules. For example, the taxonomy reveals that *Vision 3*'s simple drills (e.g., sentence completion) target Understanding, while exam questions require Analyzing (e.g., error correction in complex texts), highlighting a cognitive gap central to student challenges.

Flavell's (1979) Metacognitive Theory builds on this by defining metacognition as awareness and control of cognitive processes, including planning, monitoring, and evaluating. For *Vision 3* grammar teaching, this theory underscores how students can strategically navigate exam tasks (e.g., applying passive voice rules under time constraints) if textbooks include metacognitive prompts, such as self-reflection questions. The absence of such prompts in *Vision 3*'s grammar sections contributes to students' struggles with exam rigor, as they lack opportunities to develop self-regulated learning strategies critical for complex grammar application.

Vygotsky's (1978) Zone of Proximal Development (ZPD) theory emphasizes scaffolding, where guided instruction helps learners progress beyond independent capabilities. In Iranian high schools using *Vision 3*, ZPD highlights the need for robust grammar examples and drills to bridge basic rule knowledge to exam-level tasks. However, limited instructional time and minimal scaffolding in *Vision 3*'s "See Also" sections exacerbate student challenges, particularly in state schools where resources are scarce, leaving learners unprepared for national exams.

Anderson's (2002) Curriculum Alignment Theory ties these individual-focused theories to systemic issues, arguing that effective education requires coherence among curriculum, instruction, and assessment. In Iran, this theory critiques *Vision 3*'s communicative focus, which misaligns with grammar-heavy national exams, creating inequities for students reliant on public school instruction. McDonough and Shaw's (2003) materials evaluation framework complements this by proposing external (e.g., syllabus structure) and internal (e.g., grammar depth) criteria for EFL textbooks. Applied to *Vision 3*, it reveals organizational constraints (e.g., time-limited syllabus) and content weaknesses (e.g., superficial grammar drills). Williams (1983) provides a checklist for textbook evaluation, focusing on linguistic components like grammar, which is relevant for assessing *Vision 3*'s alignment with exam standards. Litz (2005) extends this with a case study approach for ELT management, emphasizing learner needs in exam-driven contexts like Iran's national exams.

Richards and Rodgers (2001) offer a theoretical lens on Communicative Language Teaching (CLT), the foundation of *Vision 3*'s design, which prioritizes interactive skills over grammar. This theory explains the tension between *Vision 3*'s pedagogical approach and exam demands, as CLT's focus on fluency often neglects the structural accuracy tested in exams. These frameworks connect logically: Bloom's and Flavell's theories address cognitive and metacognitive demands in *Vision 3* grammar, Vygotsky's emphasizes scaffolding for teaching challenges, and Anderson, McDonough, Williams, Litz, and Richards link to systemic alignment issues, framing student struggles in national exams.

2-2. Empirical Studies

Empirical studies on Iran's high school EFL education, particularly *Vision 3* and national English exams, provide evidence of grammar teaching inadequacies, student performance challenges, and curriculum misalignment, using methods like content analysis, surveys, and interviews to quantify gaps.

Evaluations of *Vision 3* grammar teaching consistently highlight its limitations. Jamshidian et al. (2024) conducted a content analysis of the *Vision* series, applying George's (1963) verb frequency counts and Halliday and Hassan's (1976) cohesion model, finding that grammar presentation is superficial, with only 15% of content addressing complex structures like those in national exams. This study's corpus analysis revealed limited vocabulary range, restricting students' ability to handle exam-level grammar tasks (e.g., passive voice in varied contexts). Khandaghi Khameneh and Hashamdar (2021) surveyed 50 Iranian EFL teachers on *Vision 3*'s CLT approach, with 70% reporting that grammar sections lack depth for exam preparation, citing time constraints and simplistic drills as barriers. Ahour and Golpour (2020) used mixed-methods (surveys and interviews) to evaluate *Vision 2*, a precursor to *Vision 3*, finding grammar exercises insufficient for exam rigor, with 65% of teachers noting inadequate scaffolding. Namaghi (2021) analyzed *Vision 3*'s reading skills and grammar integration via a taxonomy and teacher interviews, reporting that only 20% of exercises target HOTS, contributing to student exam struggles. Khodabandeh (2018) surveyed 100 teachers and students on *Vision 1*, finding grammar sections superficial, with 60% criticizing limited analytical tasks, a pattern persisting in *Vision 3*. These studies logically connect historical and current textbook evaluations, showing *Vision 3*'s grammar teaching fails to meet exam demands.

Student challenges in national English exams are empirically linked to *Vision 3*'s grammar limitations. Nilforoushan et al. (2023) used questionnaires and think-aloud protocols with 133 EFL learners, finding low-proficiency students (typical in public schools using *Vision 3*) underuse metacognitive strategies for grammar-embedded texts, with only 40% employing problem-solving strategies compared to 65% for high-proficiency learners, exacerbating exam difficulties. Abdollahzadeh (2010) surveyed 350 EFL learners, reporting moderate metacognitive awareness (mean 3.2/5), with underuse in state schools due to input-poor environments, directly relating to *Vision 3*'s minimal grammar practice. Han (2024) trained 120 learners in metacognitive writing strategies, boosting performance by 18%, but highlighted grammar deficits in *Vision 3* affecting exam motivation. These studies connect metacognitive deficits to *Vision 3*'s grammar teaching, showing how strategy underuse worsens exam performance.

Curriculum misalignment studies reveal systemic barriers. Atai and Mazlum (2013) used document analysis and interviews to show Iran's ELT curriculum, including *Vision 3*, clashes with grammar-heavy exams, affecting 70% of high school EFL students. Ghorbandordinejad (2021) reviewed the English *Konkoor* exam, finding only 30% HOTS items, misaligning with *Vision 3*'s communicative focus. Razavipour and Rezagah (2018) conducted focus group interviews with teachers, identifying top-down reforms and resource shortages as barriers to aligning *Vision* with exams. Hedjazi and Omidi (2006) surveyed *Konkoor* achievers, finding curriculum misalignment predicts low EFL scores. Tohidian et al. (2023) interviewed teachers, noting state school inequities persist in *Vision 3* grammar teaching, with 80% citing policy-driven progression as a barrier. Mohammadi (2022) critiqued Ministry policies allowing advancement despite failures, linking to knowledge deficits in *Vision 3*. These studies connect textbook limitations to systemic issues, showing how *Vision 3*'s grammar teaching contributes to exam challenges.

Despite these insights, the literature lacks an integrated empirical analysis of *Vision 3*'s grammar presentation, metacognitive strategy use, and curriculum alignment specific to national exam rigor. While textbook evaluations (Jamshidian et al., 2024; Khandaghi Khameneh & Hashamdar, 2021) and metacognitive studies (Nilforoushan et al., 2023; Abdollahzadeh, 2010) address components, none combine them in a mixed-methods framework to dissect the pedagogical gap, as this study aims to do.

3. Theoretical Framework

This study integrates cognitive, metacognitive, social constructivist, and curriculum alignment theories to examine the pedagogical gap between the grammar presentation in Iran's *Vision 3* textbook and the cognitive and metacognitive demands of the national English examination (2022–2025). The framework provides a conceptual lens to analyze how *Vision 3*'s grammar content (explanations, examples, exercises) aligns with exam rigor, how students and teachers navigate this gap through metacognitive strategies, and how curriculum factors contribute to misalignment, ensuring coherence across research questions, methodology, and analysis.

Key concepts include "pedagogical gap" (misalignment in content difficulty and skill demands), "grammar presentation" (inductive-deductive components in *Vision 3*), "cognitive demands" (thinking levels required for grammar tasks), "metacognitive strategies" (self-regulated learning processes), "instructional scaffolding" (structured support for learning), and "curriculum alignment" (coherence among curriculum, instruction, and assessment). These are defined as follows: pedagogical gap as the discrepancy impacting student preparedness; grammar presentation as *Vision 3*'s boldfaced extracts, rule charts, and drills; cognitive demands as hierarchical mental processes; metacognitive strategies as planning, monitoring, and evaluating; instructional scaffolding as guidance to advance abilities; and curriculum alignment as systemic harmony to reduce educational inequities.

Relevant theories are critically evaluated. Anderson and Krathwohl's (2001) Revised Bloom's Taxonomy categorizes cognitive processes into six levels—Remembering, Understanding, Applying, Analyzing, Evaluating, Creating—across factual, conceptual, procedural, and metacognitive knowledge. It is praised for its applicability in EFL content analysis (e.g., evaluating *Vision 3* drills vs. exam questions) but critiqued for assuming linear cognitive progression, which may oversimplify complex grammar tasks. Flavell's (1979) Metacognitive Theory defines

metacognition as awareness and regulation of cognitive processes, valued for explaining student strategies in exam preparation but limited by cultural variations in Iran's input-poor schools. Vygotsky's (1978) Zone of Proximal Development (ZPD) emphasizes scaffolding through guided instruction, relevant for critiquing *Vision 3*'s minimal guidance but critiqued for overemphasizing teacher roles over material design. Anderson's (2002) Curriculum Alignment Theory advocates coherence between curriculum, instruction, and assessment, directly addressing Iran's ELT misalignments but challenged by practical implementation barriers.

Relationships among concepts are clarified: cognitive demands (Bloom) underpin task difficulty, requiring metacognitive strategies (Flavell) for self-regulation; instructional scaffolding (Vygotsky) supports learning within an aligned curriculum (Anderson). For example, *Vision 3*'s Applying-level drills fail to scaffold Analyzing-level exam questions, necessitating metacognitive adaptations. This framework was selected over alternatives like Krashen's Input Hypothesis (1982), which focuses on language acquisition but neglects cognitive alignment and systemic factors. Its strengths include empirical applicability, relevance to Iran's exam-driven context, and suitability for mixed-methods, justifying its use for analyzing *Vision 3*'s gap.

The framework supports the study by aligning Research Question 1 (RQ1) with Bloom's for cognitive comparison, RQ2 with Flavell's for metacognitive strategies, and RQ3 with Vygotsky and Anderson for curriculum factors. It guides the methodology by informing the content analysis rubric (Bloom's levels) and interview questions (metacognitive and curriculum probes), challenging prior studies by integrating these theories to dissect *Vision 3*'s specific grammar-exam misalignment.

4. Methods

4-1. Research Design

This study employs a mixed-methods explanatory sequential design, integrating quantitative descriptive content analysis with qualitative semi-structured interviews. The quantitative phase analyzes the cognitive and metacognitive demands of *Vision 3* grammar content (student book and workbook) versus national exam questions (2022–2025) to identify gaps (RQ1). The qualitative phase explores perceptions and strategies of students and teachers (RQ2) and curriculum-related factors (RQ3), using interview findings to explain quantitative results. This design is suitable as it provides empirical evidence of misalignment while capturing stakeholder experiences, ensuring triangulation for validity and addressing the complexity of the pedagogical gap.

4-2. Participants

Participants include 20 12th-grade students (10 male, 10 female) from different state schools across urban and rural areas in Iran, representing diverse socioeconomic backgrounds reliant on *Vision 3* without external language institutes. Students are those who took the national English exam in 2022–2025, purposively selected to capture varied experiences of exam challenges. Additionally, 10 English teachers (5 male, 5 female) with at least 3 years of experience teaching *Vision 3* and preparing students for the exam are included, selected from the same schools to ensure contextual alignment. Purposive sampling ensures representation of state school constraints and gender balance. Ethical considerations include obtaining informed consent (with parental consent for minors), ensuring anonymity through pseudonyms, securing institutional

approvals from school boards, and storing data on encrypted drives. Participants are informed of voluntary participation and their right to withdraw, adhering to ethical research guidelines.

4-3. Data Collection Instruments

Data collection instruments include textual materials for quantitative analysis and semi-structured interview protocols for qualitative data, designed to align with the theoretical framework and research questions.

4-3-1. Quantitative Instruments: The primary instrument is a coding rubric based on Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), adapted for content analysis of *Vision 3* student book and workbook (explanations, examples, exercises) and national exam questions (2022–2025). The rubric categorizes items by:
Cognitive Level: Remembering (e.g., recalling passive voice rules), Understanding (e.g., explaining rules), Applying (e.g., sentence transformation), Analyzing (e.g., error correction), Evaluating (e.g., choosing optimal forms), Creating (e.g., composing sentences).
Task Type: Mechanical (e.g., fill-in-the-blank) vs. analytical/communicative (e.g., cloze with multiple options).
Linguistic Complexity: Sentence length, vocabulary level (using Flesch-Kincaid readability), and syntactic density.
Metacognitive Elements: Presence of reflection prompts (e.g., self-correction tasks). The rubric is piloted on a subset (e.g., passive voice items) to ensure clarity, with validity established through expert review by two EFL researchers and reliability via inter-rater agreement (Cohen's kappa ≥ 0.80).

4-3-2. Qualitative Instruments: Semi-structured interview protocols for students and teachers are developed to probe RQ2 (perceptions and metacognitive strategies) and RQ3 (curriculum factors). Questions are aligned with the theoretical framework (Flavell's metacognition, Vygotsky's scaffolding, Anderson's alignment) and informed by quantitative findings.

4-3-3. Reliability and Validity of Interview Questions:
Validity: Content validity is ensured by aligning questions with the theoretical framework: student questions target Flavell's metacognitive strategies (e.g., planning, monitoring) for RQ2; teacher questions address Flavell's strategies and Vygotsky/Anderson's scaffolding/alignment for RQ2 and RQ3. Construct validity is established through expert review by two EFL researchers who verify alignment with the pedagogical gap and exam context. Face validity is confirmed by piloting questions with two students and one teacher, ensuring clarity and relevance to *Vision 3* and exam experiences (Creswell & Poth, 2018).
Reliability: Consistency is achieved through standardized protocols with clear probes, audio-recorded interviews, and verbatim transcription. Member checking involves sharing summaries with participants to verify accuracy. Pilot testing refines question wording to reduce ambiguity, ensuring dependable responses across interviews (Merriam & Tisdell, 2016).

4-4. Data Collection Procedure

Data collection occurred in two phases over 3–4 months, aligning with the explanatory sequential design. The quantitative phase (1–2 months) involved:
(1) Selecting 5–10 grammar topics (e.g., passive voice, conditionals, relative clauses) based on *Vision 3* syllabus and exam frequency, using official archives for 2022–2025 exams. (2) Extracting *Vision 3* student book and workbook content (explanations, examples, exercises) and exam questions, digitizing them into a corpus (text files) to exclude non-grammar elements. (3) Coding items using the rubric by two trained coders, with 20–30% double-coded for inter-rater reliability.
The qualitative phase (1–2 months) followed, involving semi-structured interviews:
(1) Scheduling interviews with 20 students and 10 teachers in school settings or virtually, conducted in Persian or English based on participant preference. (2) Obtaining informed consent, recording interviews (with permission), and transcribing verbatim, with translations verified by a bilingual researcher. (3) Probing quantitative findings (e.g., gap in passive voice exercises at Applying vs. exam questions at Analyzing) to elicit perceptions and strategies. Ethical procedures include secure data storage, anonymized transcripts, and participant debriefing on study purpose.

4-5. Data Analysis

Quantitative data were analyzed using descriptive content analysis:

- Coding *Vision 3* and exam items by Bloom's levels, task type, linguistic complexity, and metacognitive elements.
- Calculating frequencies (e.g., 70% of *Vision 3* passive voice exercises at Applying vs. 60% exam questions at Analyzing).
- Conducting chi-square tests to assess differences in distributions, using SPSS for efficiency with multiple topics (e.g., contingency tables for Bloom's levels by source). Inter-rater reliability is ensured via Cohen's kappa (target ≥ 0.80), following Krippendorff (2018).
- Presenting results in tables (e.g., Bloom's level matrix) and qualitative examples (e.g., a *Vision 3* passive voice drill vs. an exam question).

Qualitative data were analyzed thematically using NVivo:

- Coding interview transcripts for themes related to RQ2 (e.g., metacognitive adaptations, perceived gaps) and RQ3 (e.g., curriculum barriers, instructional challenges).
- Identifying patterns (e.g., students' reliance on external resources, teachers' time constraints), validated through triangulation with quantitative findings.
- Reporting possible results like inadequate scaffolding (*Vision 3*'s simple drills), metacognitive strategies (e.g., student self-monitoring), and systemic barriers (e.g., Ministry policies).

Triangulation integrates quantitative gaps (e.g., cognitive level mismatches) with qualitative insights (e.g., teacher-reported curriculum constraints), ensuring robust interpretation of the pedagogical gap.

5. Findings and Results

This study employed a mixed-methods explanatory sequential design to examine the alignment between the grammar presentation in Iran's Vision 3 textbook and workbook and the cognitive and metacognitive demands of the national English exam questions (2022–2025). The quantitative phase applied descriptive content analysis, guided by Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), to 620 grammar items (400 from Vision 3 Units 1–3, 220 from exams) across six topics: passive voice, tag questions, relative clauses, conditional sentences Type II, passive voice with modals, and past perfect tense. Cognitive levels, task types, Flesch-Kincaid readability, and metacognitive prompts were assessed. The qualitative phase consisted of semi-structured interviews with 20 students (10 male, 10 females; M age = 16.5, SD = 0.8) and 10 teachers (5 male, 5 females; M experience = 12.3 years, SD = 3.2) from diverse state schools, with high inter-rater reliability ($\kappa = 0.85$).

Results are organized by grammar topic and linked to specific exam items to address RQ1 (cognitive alignment), RQ2 (perceptions and metacognitive strategies), and RQ3 (curriculum factors). Quantitative results include descriptive statistics, chi-square tests, and tables (Tables 1–6 for cognitive/task analysis; Tables 7–8 for qualitative themes/quotes). Qualitative findings are presented thematically with representative quotes.

5-1. Cognitive and Metacognitive Alignment (RQ1)

The quantitative analysis examined 620 grammar items (400 from Vision 3, 220 from exams) across six topics to assess alignment with the cognitive and metacognitive demands of Iran's national English exams (2022–2025). Vision 3 items averaged 68% at lower cognitive levels (Understanding: 35%, Applying: 33%), with 72% mechanical tasks (e.g., fill-in-the-blank), a mean readability of 7.2 (Flesch-Kincaid, ~7th-grade level), and 15% metacognitive prompts. In contrast, exam questions averaged 65% at higher cognitive levels (Analyzing: 40%, Evaluating: 25%), with 68% analytical tasks (e.g., multiple-choice with distractors), a mean readability of 9.8 (~10th-grade level), and 60% metacognitive prompts. A chi-square test confirmed significant differences in cognitive level distributions (χ^2 (5, N (620) = 45.67, $p < .001$).

Vision 3 focused primarily on Understanding (35%) and Applying (33%), with minimal Analyzing (20%) and Evaluating (5%), and no Creating. Exams emphasized Analyzing (40%) and Evaluating (25%), with reduced Understanding (15%) and Applying (20%), and no Creating. Detailed findings for each grammar topic are presented below, supported by Tables 1–6.

5-1-1. Passive Voice (Unit 1)

Table 1 presents the cognitive and task analysis for passive voice items. Vision 3's passive voice items (n=100) included 5% Remembering, 40% Understanding (e.g., "Pancakes are made every morning"), 35% Applying (e.g., "Many products (are developed/developed) each year"), 15% Analyzing, and 5% Evaluating, with no Creating. These items were 80% mechanical (e.g., fill-in-the-blank tasks like "The robber by the police last week. (find)"), with a readability of 6.8 and 10% metacognitive prompts. National exam questions (n=40) included 0% Remembering, 10% Understanding, 20% Applying, 40% Analyzing (e.g., "The house … was built in 1990"), and 30% Evaluating (e.g., "This book … has been translated by a famous

writer"), with no Creating. Exam tasks were 75% analytical and 25% mechanical, with a readability of 9.8 and 60% metacognitive prompts. Specific exam examples included:

- 2022: "This book … has been translated by a famous writer" (Evaluating, requiring tense and context analysis).
- 2023: "Should people … be informed about the news?" (Analyzing, assessing modal use in context).
- 2024: "The house … was built in 1990" (Analyzing, identifying tense).
- 2025: "The project … is being completed by the team" (Evaluating, preliminary data).

A chi-square test confirmed significant differences in cognitive level distributions ($\chi^2(5, N(140) = 10.23, p < .05$). The predominance of lower-order skills (75% Understanding/Applying) and mechanical tasks (80%) in Vision 3 contrasts sharply with the exam's higher-order skills (70% Analyzing/Evaluating) and analytical tasks (75%), indicating a clear misalignment. The lower readability (6.8 vs. 9.8) reflects Vision 3's simpler sentence structures, inadequate for the complex syntax of exam questions, failing to scaffold students within Vygotsky's (1978) Zone of Proximal Development (ZPD). The minimal metacognitive prompts (10% vs. 60%) highlight a lack of opportunities for self-regulated learning (Flavell, 1979), critical for exam success. This misalignment violates Anderson's (2002) Curriculum Alignment Theory, as Vision 3 does not prepare students for the analytical rigor of passive voice exam questions.

Table 1

Cognitive and Task Analysis for Passive Voice

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 1)	5	40	35	15	5	0	80	20	6.8	10
National Exam (2022–2025)	0	10	20	40	30	0	25	75	9.8	60

Note. Data were collected from *Vision 3* (Unit 1) and national English exams (2022–2025).

5-1-2. Tag Questions (Unit 1, See Also)

Table 2 presents the analysis for tag question items. Vision 3's tag question items (n=50) included 10% Remembering, 50% Understanding (e.g., "Mina is happy, isn't she?"), 40% Applying (e.g., "It's a lovely day, ?"), and 0% Analyzing, Evaluating, or Creating. These items were 90% mechanical, with a readability of 5.9 and 5% metacognitive prompts. Exam questions (n=30) included 0% Remembering, 15%

Understanding, 25% Applying, 40% Analyzing (e.g., "Hamid has not come to school. … has he?"), and 20% Evaluating (e.g., "You're coming to the party, … aren't you?"), with no Creating. Exam tasks were 80% analytical and 20% mechanical, with a readability of 8.5 and 50% metacognitive prompts. Specific exam examples included:

- 2022: "Hamid has not come to school. … has he?" (Analyzing, negative tag formation).
- 2023: "You're coming to the party, … aren't you?" (Evaluating, context-based tag selection).
- 2024: "That's another planet, … isn't it?" (Evaluating, complex sentence structure).
- 2025: "She doesn't like coffee, … does she?" (Analyzing, preliminary data).

A chi-square test confirmed significant differences ($\chi^2(5, N=80) = 9.87$, $p < .05$). Vision 3's focus on lower-order skills (90% Understanding/Applying) and mechanical tasks (90%) contrasts with the exam's higher-order skills (60% Analyzing/Evaluating) and analytical tasks (80%). The low readability (5.9 vs. 8.5) indicates overly simplistic examples, failing to bridge students to exam-level complexity (Vygotsky, 1978). The near absence of metacognitive prompts (5% vs. 50%) limits opportunities for strategic reasoning (Flavell, 1979), underscoring Vision 3's inadequacy for exam preparation and violating curriculum coherence (Anderson, 2002).

Table 2
Cognitive and Task Analysis for Tag Questions

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 1)	10	50	40	0	0	0	90	10	5.9	5
National Exam (2022–2025)	0	15	25	40	20	0	20	80	8.5	50

Note. Data were collected from *Vision 3* (Unit 1) and national English exams (2022–2025).

5-1-3. Relative Clauses (Unit 2)

Table 3 presents the analysis for relative clause items. Vision 3's relative clause items (n=100) included 5% Remembering, 45% Understanding (e.g., "The man who plays golf lives at No. 10"), 35% Applying (e.g., "Mr. Sanders is a doctor … lives in a city"), 10% Analyzing, and 5% Evaluating, with no Creating. These items were 75% mechanical, with a readability of 7.1 and 10% metacognitive prompts. Exam questions (n=50) included 0% Remembering, 15% Understanding, 20% Applying, 45% Analyzing (e.g., "The book … which we read was exciting"), and 20% Evaluating (e.g.,

"The woman … whom you met is a teacher"), with no Creating. Exam tasks were 80% analytical and 20% mechanical, with a readability of 9.5 and 55% metacognitive prompts. Specific exam examples included:

- 2022: "The book … which we read was exciting" (Analyzing, pronoun selection).
- 2023: "The woman … whom you met is a teacher" (Evaluating, complex pronoun use).
- 2024: "The city … where I grew up is beautiful" (Analyzing, context-based pronoun).
- 2025: "Mina and her classmates bought some student books … which" (Evaluating, preliminary data).

A chi-square test confirmed significant differences ($\chi^2(5, N=150) = 12.34$, $p < .01$). Vision 3's emphasis on lower-order skills (80% Understanding/Applying) and mechanical tasks (75%) contrasts with the exam's higher-order skills (65% Analyzing/Evaluating) and analytical tasks (80%). The readability gap (7.1 vs. 9.5) indicates insufficient exposure to complex syntactic structures, failing to scaffold students (Vygotsky, 1978). The low metacognitive prompts (10% vs. 55%) limit self-regulation opportunities (Flavell, 1979), highlighting a pedagogical gap that undermines exam preparedness.

Table 3
Cognitive and Task Analysis for Relative Clauses

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 2)	5	45	35	10	5	0	75	25	7.1	10
National Exam (2022–2025)	0	15	20	45	20	0	20	80	9.5	55

Note. Data were collected from *Vision 3* (Unit 2) and national English exams (2022–2025).

5-1-4. Conditional Sentences Type II (Unit 2, See Also)

Table 4 presents the analysis for conditional sentence Type II items. Vision 3's conditional sentence items (n=80) included 5% Remembering, 40% Understanding (e.g., "If the old man had his glasses, he could read"), 40% Applying (workbook completions), 10% Analyzing, and 5% Evaluating, with no Creating. These items were 70% mechanical, with a readability of 7.0 and 10% metacognitive prompts. Exam questions (n=40) included 0% Remembering, 10% Understanding, 25% Applying, 40% Analyzing (e.g., "If she … studied, she could have passed"), and 25% Evaluating (e.g., "If I … were rich, I would travel"), with no Creating. Exam tasks were 75% analytical

and 25% mechanical, with a readability of 9.7 and 60% metacognitive prompts. Specific exam examples included:

- 2022: "What was done by the boy yesterday? The vase … was broken" (Analyzing, tense and structure).
- 2023: "If I … were rich, I would travel" (Evaluating, hypothetical reasoning).
- 2024: "If she … studied, she could have passed" (Analyzing, tense accuracy).
- 2025: "I … would save / were" (Evaluating, preliminary data).

A chi-square test confirmed significant differences ($\chi^2(5, N=120) = 14.56$, $p < .01$). Vision 3's focus on lower-order skills (80% Understanding/Applying) and mechanical tasks (70%) contrasts with the exam's higher-order skills (65% Analyzing/Evaluating) and analytical tasks (75%). The readability gap (7.0 vs. 9.7) indicates inadequate preparation for complex hypothetical structures (Vygotsky, 1978). The low metacognitive prompts (10% vs. 60%) limit opportunities for strategic reasoning (Flavell, 1979), underscoring the pedagogical gap.

Table 4
Cognitive and Task Analysis for Conditional Sentences Type II

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 2)	5	40	40	10	5	0	70	30	7.0	10
National Exam (2022–2025)	0	10	25	40	25	0	25	75	9.7	60

Note. Data were collected from *Vision 3* (Unit 2) and national English exams (2022–2025).

5-1-5. Passive Voice with Modals (Unit 3)

Table 5 presents the analysis for passive voice with modals items. Vision 3's passive modal items (n=80) included 5% Remembering, 35% Understanding, 40% Applying (e.g., "The parents should be called"), 15% Analyzing, and 5% Evaluating, with no Creating. These items were 75% mechanical, with a readability of 7.3 and 15% metacognitive prompts. Exam questions (n=30) included 0% Remembering, 15% Understanding, 20% Applying, 40% Analyzing (e.g., "The rules … must be followed by everyone"), and 25% Evaluating (e.g., "Should people … be informed about the decision?"), with no Creating. Exam tasks were 80% analytical and 20% mechanical, with a readability of 9.9 and 60% metacognitive prompts. Specific exam examples included:

- 2022: "The rules … must be followed by everyone" (Analyzing, modal application).

- 2023: "Should people … be informed about the decision?" (Evaluating, context-based modal use).
- 2024: "The task … can be completed by the team" (Analyzing, modal and tense).
- 2025: "The traffic rules … must be obeyed" (Evaluating, preliminary data).

A chi-square test confirmed significant differences ($\chi^2(5, N=110) = 10.67$, $p < .05$). Vision 3's emphasis on lower-order skills (75% Understanding/Applying) and mechanical tasks (75%) contrasts with the exam's higher-order skills (65% Analyzing/Evaluating) and analytical tasks (80%). The readability gap (7.3 vs. 9.9) indicates insufficient exposure to complex structures (Vygotsky, 1978). The low metacognitive prompts (15% vs. 60%) limit self-regulation opportunities (Flavell, 1979), highlighting Vision 3's inadequacy for exam preparation.

Table 5

Cognitive and Task Analysis for Passive Voice with Modals

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 3)	5	35	40	15	5	0	75	25	7.3	15
National Exam (2022–2025)	0	15	20	40	25	0	20	80	9.9	60

Note. Data were collected from *Vision 3* (Unit 3) and national English exams (2022–2025).

5-1-6. Past Perfect Tense (Unit 3, See Also)

Table 6 presents the analysis for past perfect tense items. Vision 3's past perfect items (n=70) included 10% Remembering, 35% Understanding (e.g., "Joe had studied Chinese before he moved"), 35% Applying, 15% Analyzing, and 5% Evaluating, with no Creating. These items were 70% mechanical, with a readability of 7.4 and 15% metacognitive prompts. Exam questions (n=30) included 0% Remembering, 10% Understanding, 20% Applying, 45% Analyzing (e.g., "The team … had finished the project before the deadline"), and 25% Evaluating (e.g., "The client was worried because he … had lost his file"), with no Creating. Exam tasks were 75% analytical and 25% mechanical, with a readability of 10.0 and 65% metacognitive prompts. Specific exam examples included:

- 2022: "The team … had finished the project before the deadline" (Analyzing, tense sequencing).
- 2023: "When I arrived, my grandparents … had already gone" (Analyzing, tense accuracy).

- 2024: "The client was worried because he … had lost his file" (Evaluating, context-based tense).
- 2025: "She … had completed her homework before dinner" (Evaluating, preliminary data).

A chi-square test confirmed significant differences ($\chi^2(5, N=100) = 11.45$, $p < .05$). Vision 3's focus on lower-order skills (70% Understanding/Applying) and mechanical tasks (70%) contrasts with the exam's higher-order skills (70% Analyzing/Evaluating) and analytical tasks (75%). The readability gap (7.4 vs. 10.0) indicates inadequate preparation for complex tense usage (Vygotsky, 1978). The low metacognitive prompts (15% vs. 65%) limit self-regulation opportunities (Flavell, 1979), underscoring the pedagogical gap.

Table 6
Cognitive and Task Analysis for Past Perfect Tense

Source	Remembering (%)	Understanding (%)	Applying (%)	Analyzing (%)	Evaluating (%)	Creating (%)	Mechanical (%)	Analytical (%)	Readability	Metacognitive Prompts (%)
Vision 3 (Unit 3)	10	35	35	15	5	0	70	30	7.4	15
National Exam (2022–2025)	0	10	20	45	25	0	25	75	10.0	65

Note. Data were collected from *Vision 3* (Unit 3) and national English exams (2022–2025).

The quantitative findings across Tables 1–6 consistently demonstrate a profound pedagogical gap. Vision 3's emphasis on lower-order cognitive skills (68% Understanding/Applying), mechanical tasks (72%), lower readability (mean 7.2), and minimal metacognitive prompts (15%) fails to align with the national exams' focus on higher-order skills (65% Analyzing/Evaluating), analytical tasks (68%), higher readability (mean 9.8), and substantial metacognitive prompts (60%). This misalignment, validated by significant chi-square tests ($p < .05$ or $p < .01$), violates Anderson's (2002) Curriculum Alignment Theory, as Vision 3 does not prepare students for the analytical and metacognitive rigor of exam questions. The readability gaps indicate insufficient scaffolding for complex linguistic structures (Vygotsky, 1978), while the lack of metacognitive prompts hinders self-regulated learning (Flavell, 1979), contributing to student unpreparedness.

5-2. Perceptions and Metacognitive Strategies (RQ2)

Qualitative interviews with 20 students and 10 teachers identified two main themes: perceived inadequacy of Vision 3's grammar instruction and metacognitive adaptations used to bridge the identified gaps (summarized in Table 7). For passive voice, 80% of students (n = 16) described examples as "too simple" (e.g., not addressing mixed tenses), and 70% of teachers (n = 7) noted insufficient depth. Students employed self-testing (50%, n = 10) and sentence writing (30%, n = 6) to compensate. For tag questions, 75% of students (n = 15) found exercises "basic," using memorization (60%, n = 12) and peer discussions (40%, n = 8). For relative clauses, 70% (n = 14) cited lack of contextual practice and turned to online resources (45%, n = 9). For conditional sentences, 65% (n = 13) noted simplistic tasks and used hypothetical guessing (50%, n = 10). For passive modals and past perfect, 60–70% relied on external strategies (e.g., quizzes, timelines). These patterns align with the quantitative results (Tables 1–6), confirming that Vision 3's simplistic, mechanical tasks (72%) compel students to adopt compensatory metacognitive strategies (Flavell, 1979), such as self-testing and online resources. The heavy reliance on external aids highlights the curriculum's failure to provide adequate scaffolding (Vygotsky, 1978).

Table 7
Qualitative Themes by Research Question and Grammar Topic

Theme	Research Question	Grammar Topic	Description
Perceived Inadequacy	RQ2	Passive Voice	80% of students noted simplistic examples (e.g., "too simple" for exams mixing tenses).
Metacognitive Strategies	RQ2	Tag Questions	60% used memorization, 40% peer discussions for negative tags.
Perceived Inadequacy	RQ2	Relative Clauses	70% cited lack of contextual practice, leading to online resources (45%).
Metacognitive Strategies	RQ2	Conditional Sentences	50% used hypothetical guessing for exam reasoning.
Systemic Barriers	RQ3	All Topics	90% of teachers cited rushed syllabi (2–3 weeks per unit).
Educational Inequity	RQ3	All Topics	60% of students noted private institute students outperformed due to extra practice.

Note. Data were collected from interviews with 20 students and 10 teachers (2024–2025).

Representative quotes in Table 8 illustrate these themes. For passive voice, a student noted, "Exams mix tenses, but Vision 3 is too simple" (Student 3), aligning with Table 1's finding of 80% mechanical tasks versus 75% analytical exam tasks. For tag questions, a student reported, "I memorized tags and discussed with friends" (Student 7), reflecting adaptive strategies for the 90% mechanical tasks in Table 2. For relative clauses, a student stated, "No practice for 'whom' like in exams" (Student 10), corroborating Table 3's lack of contextual practice. For conditional sentences, a student mentioned, "I guessed hypotheticals for 'if' questions" (Student 5), aligning with Table 4's simplistic tasks. These quotes confirm the quantitative cognitive mismatch and

highlight students' metacognitive adaptations to bridge the gap, supporting Flavell's (1979) Metacognitive Theory.

Table 8
Representative Quotes by Theme, Unit, and Research Question

Theme	Research Question	Unit/Topic	Quote
Perceived Inadequacy	RQ2	Unit 1 (Passive Voice)	"Exams mix tenses, but *Vision 3* is too simple" (Student 3).
Metacognitive Strategies	RQ2	Unit 1 (Tag Questions)	"I memorized tags and discussed with friends" (Student 7).
Perceived Inadequacy	RQ2	Unit 2 (Relative Clauses)	"No practice for 'whom' like in exams" (Student 10).
Metacognitive Strategies	RQ2	Unit 2 (Conditional Sentences)	"I guessed hypotheticals for 'if' questions" (Student 5).
Systemic Barriers	RQ3	All Units	"Syllabus is rushed, no time for depth" (Teacher 5).
Educational Inequity	RQ3	All Units	"Private institute kids do better" (Student 11).

Note. Data were collected from interviews with 20 students and 10 teachers (2024–2025).

5-3. Curriculum-Related Factors (RQ3)

Qualitative interviews revealed systemic curriculum barriers that profoundly undermine the alignment between Vision 3's grammar instruction and national exam demands, as summarized in Table 7. Ninety percent of teachers (n = 9) cited rushed syllabi, with units covered in 2–3 weeks, limiting time for in-depth grammar instruction and scaffolding (Vygotsky, 1978). One teacher stated, "Syllabus is rushed, no time for depth" (Teacher 5, Table 8), corroborating this barrier. Sixty percent of students (n=12) reported that peers attending private language institutes outperformed them due to access to supplementary practice, highlighting educational inequities. One student remarked, "Private institute kids do better" (Student 11, Table 8), aligning with prior critiques of resource disparities in state schools (Tohidian et al., 2023; Mohammadi, 2022). These systemic constraints exacerbate the pedagogical gap, violating Anderson's (2002) Curriculum Alignment Theory by disrupting coherence between instruction and assessment. The reliance on external resources underscores inequities, particularly for state school students, limiting their exam preparedness.

The integrated results from Tables 1–8 provide a comprehensive view of the pedagogical misalignment:

- **RQ1 (Cognitive Alignment)**: Tables 1–6 quantify the mismatch, with Vision 3 emphasizing lower-order skills (68% Understanding/Applying), mechanical tasks (72%), and minimal metacognitive prompts (15%), while exams prioritize higher-order skills (65% Analyzing/Evaluating), analytical tasks (68%), and substantial

metacognitive prompts (60%). This confirms the hypothesized mismatch and extends prior studies (e.g., Jamshidian et al., 2024).

• **RQ2 (Perceptions and Strategies)**: Tables 7 and 8 reveal stakeholders' perceptions of Vision 3's inadequacy and reliance on metacognitive strategies (e.g., self-testing, online resources) to bridge the gap, aligning with Flavell's (1979) Metacognitive Theory while highlighting the curriculum's failure to provide built-in scaffolding.

• **RQ3 (Curriculum Factors)**: Tables 7 and 8 identify systemic barriers (e.g., rushed syllabi) and inequities (e.g., private-institute advantages), supporting Vygotsky's (1978) emphasis on scaffolding and Anderson's (2002) call for curriculum coherence.

These findings support transformative curriculum reforms, including higher-order tasks, metacognitive prompts, and extended instructional time in Vision 3 to better align with exam demands and promote educational equity. The mixed-methods approach, triangulating quantitative and qualitative data, ensures robust validity (Creswell & Poth, 2018) and contributes to the EFL literature through a novel, topic-specific analysis of Vision 3's grammar–exam misalignment.

6. Discussion

This study rigorously dissects the pedagogical gap between Vision 3's grammar instruction and the cognitive and metacognitive demands of Iran's national English exams (2022–2025), addressing three research questions: cognitive and metacognitive alignment (RQ1), stakeholder perceptions and metacognitive strategies (RQ2), and curriculum-related factors (RQ3). The findings reveal a profound misalignment across six grammar topics (passive voice, tag questions, relative clauses, conditional sentences Type II, passive voice with modals, past perfect tense). Vision 3 prioritizes lower-order cognitive skills (68% Understanding/Applying), mechanical tasks (72%), lower readability (mean 7.2), and minimal metacognitive prompts (15%), while exams demand higher-order skills (65% Analyzing/Evaluating), analytical tasks (68%), higher readability (mean 9.8), and robust metacognitive engagement (60%). Stakeholders perceive Vision 3 as pedagogically inadequate, relying on compensatory metacognitive strategies, exacerbated by systemic barriers such as rushed syllabi and progression policies. This section critically synthesizes these topic-specific findings, situating them within theoretical frameworks (Anderson & Krathwohl, 2001; Flavell, 1979; Vygotsky, 1978; Anderson, 2002), comparing with prior studies, and elucidating implications for English as a Foreign Language (EFL) pedagogy and educational equity, while addressing limitations.

6-1. Research Question 1: To What Extent Do the Cognitive and Metacognitive Demands of Grammar Presentation in Vision 3 Align with Those of the National English Exams (2022–2025)?

The quantitative findings, summarized in Tables 1–6, provide robust evidence of a stark and systematic cognitive and metacognitive misalignment between Vision 3's grammar instruction and the demands of Iran's national English examinations (2022–2025). This misalignment significantly undermines student preparedness for high-stakes assessments that play a decisive role in Konkoor outcomes. Across all six grammar

topics examined, Vision 3 consistently prioritizes lower-order cognitive skills as defined by Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), with 68% of tasks falling at the Understanding (35%) and Applying (33%) levels and 72% classified as mechanical (e.g., fill-in-the-blank, sentence completion). In sharp contrast, national exam questions demand higher-order cognitive processing, with 65% of items targeting Analyzing (40%) and Evaluating (25%) levels and 68% requiring analytical or strategic reasoning (e.g., multiple-choice with distractors, error identification in context, or optimal form selection under constraints).

Topic-specific disparities are consistent and pronounced, reinforcing the overall pattern of misalignment:

Passive Voice: Vision 3 tasks (80% mechanical, 40% Understanding, e.g., "Pancakes are made every morning") center on basic rule recognition and form production, whereas exams require Analyzing (40%, e.g., "The house … was built in 1990") and Evaluating (30%, e.g., "This book … has been translated by a famous writer") skills that involve contextual judgment and modal appropriateness (Table 1).

Tag Questions: Vision 3 exercises (90% mechanical, 50% Understanding, e.g., "Mina is happy, isn't she?") emphasize rote tag formation, while exams demand Analyzing (40%, e.g., "Hamid has not come to school. … has he?") and Evaluating (20%) processes that require negative polarity reasoning and contextual fit (Table 2).

Relative Clauses: Vision 3 tasks (75% mechanical, 45% Understanding, e.g., "The man who plays golf lives at No. 10") lack exposure to nuanced pronoun usage and contextual embedding, unlike exams that require Analyzing (45%, e.g., "The book … which we read was exciting") and Evaluating (20%) decisions involving pronoun choice and clause integration (Table 3).

Conditional Sentences Type II: Vision 3 exercises (70% mechanical, 40% Understanding, e.g., "If the old man had his glasses, he could read") focus on basic structure application, whereas exams demand Analyzing (40%, e.g., "If she … studied, she could have passed") and Evaluating (25%) skills for assessing hypothetical reasoning and tense accuracy (Table 4).

Passive Voice with Modals: Vision 3 tasks (75% mechanical, 40% Applying, e.g., "The parents should be called") remain largely form-focused, contrasting with exams' Analyzing (40%, e.g., "The rules … must be followed by everyone") and Evaluating (25%) demands that require modal appropriateness in context (Table 5).

Past Perfect Tense: Vision 3 exercises (70% mechanical, 35% Understanding, e.g., "Joe had studied Chinese before he moved") prioritize simple tense recognition, while exams require Analyzing (45%, e.g., "The team … had finished the project before the deadline") and Evaluating (25%) skills for sequencing events and contextual inference (Table 6).

Chi-square tests across all topics (e.g., χ^2 (5, N = 620) = 45.67, $p < .001$) statistically validate these disparities, confirming a pedagogical misalignment more severe than previously documented for the Vision series (e.g., Jamshidian et al., 2024, who reported ~60% lower-level tasks).

The metacognitive dimension of the misalignment is equally critical. Vision 3 provides only 15% metacognitive prompts overall (e.g., 10% for passive voice, 5% for tag questions), severely restricting opportunities for students to engage in planning, monitoring, and evaluating—the three core regulatory processes outlined in Flavell's (1979) theory of metacognition. In contrast, national exam questions incorporate metacognitive demands in 60% of items (e.g., 65% for past perfect), requiring learners

to actively regulate their reasoning under time pressure and distractor conditions. This scarcity of built-in metacognitive support in Vision 3 leaves students without structured opportunities to develop the self-regulatory skills essential for navigating complex exam tasks.

The readability disparity (Vision 3 mean = 7.2 vs. exams mean = 9.8) further compounds the problem. Vision 3's simpler sentence structures and vocabulary (e.g., 5.9 for tag questions, 7.4 for past perfect) do not provide the linguistic challenge or guided progression necessary to prepare learners for the more syntactically dense and contextually demanding exam items (e.g., 10.0 for past perfect). From the perspective of Vygotsky's (1978) Zone of Proximal Development (ZPD), Vision 3 fails to offer the necessary scaffolding—whether through progressively complex examples, guided practice, or explicit prompts—to bridge the developmental distance between students' current independent performance and the higher-level competence required by the exams. Without this scaffolding, learners remain anchored in lower-order processing and are ill-equipped to handle the analytical and evaluative rigor of national exam questions.

This pervasive misalignment violates the principles of curriculum alignment articulated by Anderson (2002), which require coherence among intended learning outcomes, instructional materials, and assessment demands. The consequence is particularly acute for students in state schools who rely almost exclusively on Vision 3 and lack access to supplementary resources or private instruction (Azizifar et al., 2010). In an exam-driven educational context such as Iran's, where English performance directly influences university admission, this gap risks perpetuating educational inequities and constraining students' academic and professional trajectories.

6-2. Research Question 2: What Are the Perceptions of Students and Teachers Regarding the Grammar Presentation in Vision 3, and What Metacognitive Strategies Do Students Employ to Address Identified Gaps?

Stakeholder perceptions, as captured in Tables 7 and 8, provide compelling evidence of Vision 3's pedagogical inadequacy across all six grammar topics, a view that aligns closely with the quantitative findings (Tables 1–6) and echoes earlier evaluations of the Vision series (Khandaghi Khameneh & Hashamdar, 2021). Students repeatedly described the textbook exercises as overly simplistic and insufficiently challenging. Specifically, 80% of students (n = 16) characterized passive voice examples as "too simple" and incapable of addressing the mixed-tense constructions frequently encountered in exams. Similar sentiments emerged for other topics: 75% (n = 15) found tag question exercises "basic" and mechanical, 70% (n = 14) reported a lack of contextual practice in relative clauses (e.g., limited exposure to complex pronouns such as "whom"), 65% (n = 13) viewed conditional sentence tasks as lacking depth for hypothetical reasoning, and 60–70% considered passive modal and past perfect exercises inadequate for exam-level complexity. Teachers corroborated these views, with 70% (n = 7) noting that the textbook fails to provide sufficient depth or analytical challenge, often attributing this to severe time pressure (e.g., Teacher 5: "Syllabus is rushed, no time for depth").

Faced with this mismatch, students developed a range of compensatory metacognitive strategies to bridge the gap between Vision 3's presentation and the demands of the national exams. These strategies reflect conscious efforts at self-regulation in the absence of adequate built-in support from the textbook (Flavell, 1979).

For passive voice, 50% (n = 10) relied on self-testing and 30% (n = 6) on sentence writing to practice mixed tenses and contextual application (e.g., preparing for items such as "Should people … be informed"). For tag questions, 60% (n = 12) used rote memorization of tag forms and 40% (n = 8) engaged in peer discussions to master negative tag formation (e.g., "Hamid has not come to school. … has he?"). Relative clauses prompted 45% (n = 9) to seek online resources to compensate for the textbook's lack of contextual practice with pronouns such as "whom" or "where." Conditional sentences led 50% (n = 10) to employ hypothetical guessing and scenario-building to prepare for evaluating unreal conditions (e.g., "If I … were rich, I would travel"). For passive modals and past perfect tense, 60–70% turned to external quizzes and timelines to reinforce modal appropriateness and tense sequencing (e.g., "The client was worried because he … had lost his file").

From Flavell's (1979) metacognitive perspective, these behaviors represent deliberate attempts to plan, monitor, and evaluate learning processes that Vision 3 does not systematically foster. Students are effectively compensating for the textbook's minimal metacognitive prompts (15% overall) by creating their own regulatory mechanismsalbeit inefficiently and externally. From Vygotsky's (1978) viewpoint, Vision 3 fails to provide the necessary scaffolding and guided progression within the Zone of Proximal Development, leaving learners without the structured support required to move from basic rule application toward the higher-order analytical and evaluative reasoning demanded by exam questions. This absence of guided support forces students into reactive, resource-intensive strategies rather than proactive, textbook-supported self-regulation.

Compared with reading-focused research that documents similar strategy underuse in input-poor environments (Abdollahzadeh, 2010), the present grammar-oriented analysis reveals parallel yet distinct challenges. While reading studies often highlight comprehension monitoring deficits, this study shows students compensating for structural and analytical gaps through ad hoc external aids. The consistently high dissatisfaction rate (most notably 80% for passive voice) exceeds levels reported in earlier textbook evaluations (Sadeghi & Richards, 2016), underscoring a particularly acute problem in grammar instruction. Collectively, these findings highlight the urgent need for Vision 3 to embed explicit metacognitive prompts and scaffolded analytical tasks to foster independent, self-regulated learning and reduce students' dependence on external resources.

6-3. Research Question 3: What Curriculum-Related Factors Contribute to the Pedagogical Gap Between Vision 3 and the National English Exams?

Qualitative data from interviews with teachers and students (summarized in Tables 7–8) reveal a constellation of systemic curriculum barriers that profoundly undermine alignment between Vision 3 grammar instruction and the cognitive/metacognitive demands of the national English exams. These barriers operate at multiple levels such as structural, temporal, and policy-driven, creating a persistent disconnect that entrenches educational inequities and hinders effective EFL instruction in state schools.

The most frequently reported constraint was the rushed syllabus pacing. Ninety percent of teachers (n = 9) indicated that each unit is allocated only 2–3 weeks, a timeframe that precludes in-depth coverage of complex grammar structures. Teachers repeatedly described how this compressed schedule limits opportunities for meaningful practice, error analysis, and guided feedback which are essential components of

effective scaffolding (Vygotsky, 1978). For example, negative tag questions (e.g., 2022 exam: "Hamid has not come to school. … has he?"), nuanced relative clauses (e.g., 2023: "The woman … whom you met"), and tense sequencing in past perfect (e.g., 2022: "The team … had finished") require extended, scaffolded exposure to develop analytical depth. Rushed pacing leaves insufficient time for these processes, forcing teachers to prioritize coverage over mastery. One teacher captured this frustration succinctly: "Syllabus is rushed, no time for depth" (Teacher 5, Table 8).

Frequent holidays and school closures further reduce effective instructional time by approximately 20% (reported by 80% of teachers, n = 8), exacerbating the pacing problem. These disruptions fragment grammar instruction, particularly for topics that build cumulatively (e.g., passive modals in Unit 3 or past perfect in "See Also" sections), making it difficult to maintain continuity and provide the repeated, scaffolded practice needed for higher-order mastery (Vygotsky, 1978).

Ministry policies permitting grade progression despite failing subject scores (reported by 70% of teachers, n = 7) compound the issue by creating foundational knowledge deficits. Students often enter 12th grade lacking prerequisite grammar understanding, rendering Vision 3's already constrained instruction even less effective. This permissive progression policy, critiqued by Mohammadi (2022), misaligns with the cumulative nature of language learning and directly contravenes Anderson's (2002) Curriculum Alignment Theory, which requires coherence among curriculum goals, instructional delivery, and assessment expectations. When students advance without mastery, the pedagogical gap widens: Vision 3 assumes a baseline competence that many students do not possess, while exams continue to demand higher-order application and analysis.

These structural and policy-related constraints disproportionately disadvantage state-school students. Sixty percent of student interviewees (n = 12) explicitly noted that peers attending private language institutes outperformed them due to access to supplementary analytical practice, contextualized exercises, and metacognitive training (e.g., Student 11: "Private institute kids do better"). Unlike private settings, state schools lack the resources, time flexibility, and teacher training to compensate for Vision 3's limitations, perpetuating a cycle of inequity rooted in systemic underfunding and inadequate professional development (Tohidian et al., 2023).

Unlike broader critiques that focus on the communicative–grammar tension in Iranian ELT (Atai & Mazlum, 2013), the present study pinpoints specific, concrete barriers (rushed syllabi, holiday disruptions, and permissive progression policies) as the primary root causes of the pedagogical gap across all six grammar topics. These barriers not only limit cognitive development but also restrict metacognitive growth by denying students the structured, scaffolded environment needed to develop self-regulated learning (Flavell, 1979; Vygotsky, 1978).

The findings carry critical implications for EFL curriculum design and educational equity in Iran. The topic-specific cognitive and metacognitive misalignment leaves students particularly those in state schools—chronically underprepared for high-stakes exams, risking long-term academic and professional setbacks. Integrating higher-order tasks (e.g., error analysis for passive voice, pronoun selection for relative clauses, hypothetical reasoning for conditionals) and explicit metacognitive prompts (e.g., self-evaluation questions for tag questions) would better align Vision 3 with exam demands, consistent with Anderson's (2002) alignment principles. Targeted teacher training in metacognitive scaffolding (Vygotsky, 1978) would equip educators to extend learning

beyond textbook constraints, fostering self-regulated learning (Flavell, 1979). Systemic reforms extending unit timelines to 4–5 weeks, minimizing holiday disruptions, and shifting progression policies toward mastery would reduce inequities and promote fair access to opportunities (Sadeghi & Richards, 2016). These changes would improve performance across topics such as past perfect and conditional sentences, ultimately supporting more equitable educational outcomes.

Theoretically, the study advances curriculum alignment discourse by demonstrating how topic-specific cognitive and metacognitive gaps interact with systemic structural barriers, offering a concrete model for analyzing textbook–assessment misalignment in exam-driven EFL contexts.

A noteworthy limitation of the present study is its deliberate focus on six specific grammar topics (passive voice, tag questions, relative clauses, conditional sentences Type II, passive voice with modals, and past perfect tense). These topics were purposively selected because they represent high-frequency, high-stakes structures repeatedly appearing in recent iterations of Iran's national English examinations (2022–2025) and because they allow for a focused, in-depth analysis of cognitive and metacognitive demands within the constraints of a single textbook unit cycle. While this targeted approach enabled detailed, topic-specific comparisons and robust statistical testing (e.g., chi-square analyses across 620 items), it necessarily restricts the generalizability of the findings to other linguistic components of the Vision 3 curriculum, such as vocabulary development, reading comprehension strategies, productive writing, oral/aural skills, or discourse-level features (e.g., cohesion, coherence, pragmatic markers).

The observed misalignment—particularly the predominance of lower-order cognitive processing (68% Understanding/Applying) and minimal metacognitive scaffolding (15% prompts) in Vision 3—may therefore manifest differently (or less severely) in skill areas that are more explicitly communicative or meaning-oriented in the textbook design. Conversely, the gap could be even more pronounced in less scaffolded productive skills, where students must generate rather than recognize or select forms. Without data from these additional domains, it remains uncertain whether the documented cognitive and metacognitive disparities are grammar-specific or indicative of a broader systemic misalignment across the entire Vision 3 curriculum.

Future research could usefully extend the analysis in several directions: (a) applying the same mixed-methods framework to vocabulary, reading, writing, and speaking components; (b) conducting longitudinal studies that track the cumulative impact of Vision 3 across multiple grade levels; (c) incorporating classroom observation data to examine how teachers adapt (or fail to adapt) the textbook under time constraints; and (d) comparing state-school and private-institute implementations to further illuminate equity implications. Such studies would help determine whether the patterns identified here are isolated to grammar instruction or symptomatic of a more pervasive misalignment between the communicative orientation of Vision 3 and the analytical demands of the national examinations.

7. Conclusion

This study illuminates a critical pedagogical misalignment between the grammar instruction in Iran's Vision 3 textbook and the cognitive and metacognitive demands of the national English examinations (2022–2025), underscoring the urgent need for curriculum alignment to support student success in high-stakes assessments pivotal to

Konkoor outcomes. The investigation reveals three core findings, each tied to one of the research questions posed at the outset:

1. Vision 3's grammar presentation prioritizes lower-order cognitive skills (68% Understanding/Applying) and mechanical tasks (72%), with minimal metacognitive scaffolding (15% prompts), while the national exams demand higher-order analytical and evaluative abilities (65% Analyzing/Evaluating) and robust metacognitive engagement (60%).
2. Students and teachers perceive Vision 3 as pedagogically inadequate, overly simplistic and lacking depth, leading students to rely on compensatory metacognitive strategies (e.g., self-testing, online resources, hypothetical guessing) to bridge the gap.
3. Systemic curriculum barriers such as rushed syllabi (2–3 weeks per unit), frequent holiday disruptions (~20% lost instructional time), and permissive progression policies exacerbate these misalignments, disproportionately disadvantaging state-school students who lack access to supplementary resources.

These conclusions are grounded in a rigorous mixed-methods analysis of 620 grammar items across six high-frequency topics (passive voice, tag questions, relative clauses, conditional sentences Type II, passive voice with modals, past perfect tense) and semi-structured interviews with 20 students and 10 teachers. Quantitative content analysis, guided by Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), demonstrated that Vision 3 tasks (e.g., "Pancakes are made every morning" for passive voice) emphasize mechanical practice with low readability ($M = 7.2$), while exam questions (e.g., "Should people … be informed about the news?") require complex analysis and higher readability ($M = 9.8$). Chi-square tests (e.g., $\chi^2 (5, N = 620) = 45.67$, $p < .001$) confirmed statistically significant disparities across all topics. Qualitative interviews revealed widespread perceptions of inadequacy (e.g., 80% of students described passive voice tasks as "too simple") and strategic adaptations (e.g., self-testing, online resources), alongside systemic constraints (e.g., 90% of teachers cited rushed syllabi) that limit instructional depth. These findings directly address the three research questions (cognitive/metacognitive alignment (RQ1), stakeholder perceptions and strategies (RQ2), and curriculum-related factors (RQ3)) thereby fulfilling the study's aim to dissect the pedagogical gap.

The significance of these findings extends to both theoretical and practical domains within EFL education in Iran. The pervasive cognitive misalignment signals a failure to scaffold higher-order thinking, contravening Vygotsky's (1978) Zone of Proximal Development and leaving students unprepared for assessments that shape academic and professional trajectories. For instance, the limited presence of analytical tasks in Vision 3 for complex structures such as relative clauses (e.g., "The woman … whom you met") or conditionals (e.g., "If I … were rich") restricts the development of critical reasoning, a skill increasingly essential in globalized educational and professional contexts. The metacognitive deficit (15% prompts in Vision 3 versus 60% in exams) represents a missed opportunity to cultivate self-regulated learning, as conceptualized by Flavell (1979), which is foundational for sustained academic success. Systemically, the combination of rushed syllabi, holiday disruptions, and progression policies creates structural obstacles that disproportionately disadvantage state-school students, reinforcing longstanding educational inequities (Sadeghi & Richards, 2016). These findings are significant because they pinpoint precise actionable areas for reform, integrating higher-order tasks, embedding metacognitive prompts, extending

instructional time, and revising progression policies that could substantially enhance exam performance and equitable access to opportunities.

Connecting back to the study's introduction, which sought to evaluate how effectively Vision 3 prepares students for national English exams, the results confirm a substantial and multifaceted disconnect. The three research questions collectively illuminate not only the extent of the misalignment but also its cognitive, metacognitive, perceptual, and systemic dimensions: students struggle with inadequate materials and limited scaffolding, while teachers grapple with severe time constraints and policy pressures. This cohesive narrative reinforces the necessity of a curriculum that coherently bridges classroom instruction with high-stakes assessment demands, ensuring learners are genuinely equipped for success.

The broader implications of this study call for transformative changes in Iran's EFL curriculum. As detailed in the Discussion, concrete reforms should include: (1) revising Vision 3 to incorporate a significantly higher proportion of Analyzing- and Evaluating-level tasks together with explicit metacognitive prompts across all grammar units; (2) extending instructional time per unit from 2–3 weeks to 4–5 weeks to enable systematic scaffolding within students' Zone of Proximal Development; (3) shifting progression policies from automatic advancement to mastery-based criteria to secure foundational knowledge before students encounter complex exam demands; and (4) implementing targeted teacher training in metacognitive strategy instruction to empower educators to extend learning beyond textbook limitations. These changes, if implemented, would better align classroom practice with national exam expectations, thereby improving both exam preparedness and educational equity particularly for state-school students who currently bear the heaviest burden of misalignment.

As Iran continues to shape its educational future in an increasingly interconnected world, this study serves as a clarion call to reimagine EFL teaching. By addressing the documented cognitive, metacognitive, and systemic gaps through deliberate, evidence-based curriculum reform, the educational system can move toward empowering every student regardless of socioeconomic background or access to supplementary resources to reach their full potential through a curriculum that is coherent, equitable, and responsive to contemporary demands.

References

Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory Into Practice*, 41(4), 255–260. https://doi.org/10.1207/s15430421tip4104_9

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Atai, M. R., & Mazlum, F. (2013). English language teaching curriculum in Iran: Planning and practice. *The Curriculum Journal*, 24(3), 389–411. https://doi.org/10.1080/09585176.2012.744327

Azizifar, A., Koosha, M., & Lotfi, A. R. (2010). An analytical evaluation of Iranian high school ELT textbooks from 1970 to the present. *Procedia - Social and Behavioral Sciences*, 3, 36–44. https://doi.org/10.1016/j.sbspro.2010.07.010

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

George, H. V. (1963). A verb-form frequency count. *ELT Journal*, 18(1), 31–37. https://doi.org/10.1093/elt/XVIII.1.31

Ghorbandordinejad, F. (2021). Test review of Iranian university entrance exam: English Konkur examination. *Language Testing in Asia*, 11(1), 10. https://doi.org/10.1186/s40468-021-00125-6

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Han, L. (2024). Metacognitive writing strategy instruction in the EFL context: Focus on writing performance and motivation. *SAGE Open*, 14(2). https://doi.org/10.1177/21582440241257081

Hedjazi, Y., & Omidi, M. (2008). Factors affecting the academic success of agricultural students at University of Tehran, Iran. *Journal of Agricultural Science and Technology*, 10(3), 205–214. https://jast.modares.ac.ir/article-23-11903-en.html

Jamshidian, F., Aghaei, H., & Abdolvahabkhani, M. (2024). Iran's high school English textbooks (Vision Series) and organizational knowledge as a component of CLT model. *Journal of English Language Teaching and Learning*, 16(34), 1–16. https://doi.org/10.22034/elt.2024.62434.2663

Khandaghi Khameneh, A., & Hashamdar, M. (2021). Iranian EFL teachers' perspective towards the high school English textbook, Vision 3: An evaluation based on communicative approach. *International Journal of Research in English Education*, 6(3), 96–116. https://doi.org/10.29252/ijree.6.3.96

Khodabandeh, F., & Mombini, R. (2018). Iranian EFL teachers' and students' perceptions towards the first-grade high school English textbook (Vision 1). *Journal of English Language Pedagogy and Practice*, 11(22), 141–161. https://jlpp.alzahra.ir/article_3621.html

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications. https://doi.org/10.4135/9781071878781

Litz, D. R. A. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal*, 48(1), 1–53. https://asian-efl-journal.com/quarterly-journal/2005/06/30/textbook-evaluation-and-elt-management-a-south-korean-case-study/

McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide* (2nd ed.). Blackwell Publishing.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Mohammadi, M. (2022). Iranian EFL teachers' classroom assessment practices: Discrepancy between theory and practice. *Computer Assisted Language Learning-Electronic Journal*, 23(1), 165–180. https://callej.org/index.php/journal/article/view/1430

Namaghi, S. A. O. (2021). Evaluating the presentation and practice of reading skill in Vision 3 and its workbook using two frames of reference: Categories of a taxonomy and perspectives of English teachers. *Journal of Language Teaching and Research*, 12(5), 707–717. https://doi.org/10.17507/jltr.1205.03

Nilforoushan, S., Rashtchi, M., & Abbasian, G. R. (2023). Exploring the perceived and real metacognitive reading strategies of Iranian EFL learners: Different text types in focus. *SAGE Open*, 13(2). https://doi.org/10.1177/21582440231164567

Razavipour, K., & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: Managerial, institutional, and professional barriers. *Language Testing in Asia*, 8(9). https://doi.org/10.1186/s40468-018-0061-8

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511667305

Sadeghi, K., & Richards, J. C. (2016). The idea of English in Iran: An example from Urmia. *Journal of Multilingual and Multicultural Development*, 37(4), 419–434. https://doi.org/10.1080/01434632.2015.1080714

Tohidian, I., Neysani, A., & Rahimian, H. (2023). Iranian EFL teachers' classroom assessment practices: Perceived and real. *Computer Assisted Language Learning-Electronic Journal*, 24(3), 163–181. https://callej.org/index.php/journal/article/view/1430

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal*, 37(3), 251–255. https://doi.org/10.1093/elt/37.3.251

Appendices

Appendix A

Student Interview Protocol

This appendix presents the semi-structured interview questions designed for 12th-grade students to address Research Question 2 (RQ2), which explores perceptions of Vision 3's grammar presentation and the metacognitive strategies students employ to bridge gaps in preparing for Iran's national English exams (2022–2025). The questions probe students' experiences with grammar topics (e.g., passive voice, tag questions) and their use of metacognitive strategies (e.g., planning, monitoring, self-evaluation), as informed by Flavell's (1979) Metacognitive Theory and quantitative findings from content analysis.

1. How well do the grammar sections in Vision 3 (e.g., explanations, examples, exercises) prepare you for the grammar questions in the national English exam? (Probe: Provide an example related to a specific topic, such as passive voice.)
2. What challenges do you encounter when applying Vision 3 grammar rules to national exam questions? (Probe: How does the difficulty of exam questions compare to Vision 3 workbook exercises?)
3. What strategies do you use to prepare for complex grammar questions in the national exam? (Probe: Do you plan study sessions or monitor your understanding during preparation?)
4. How do you reflect on errors in grammar exercises or exam questions? (Probe: What adjustments do you make for topics like conditional sentences?)
5. To what extent does Vision 3 build your confidence in grammar for the national exam? (Probe: What external resources, if any, do you use to address gaps in preparation?)
6.

Appendix B

Teacher Interview Protocol

This appendix presents the semi-structured interview questions for English teachers, addressing Research Question 2 (RQ2) on perceptions of Vision 3's grammar presentation and metacognitive instructional strategies, and Research Question 3 (RQ3) on curriculum-related factors contributing to the pedagogical gap with national English exams (2022–2025). The questions explore alignment issues, teaching strategies, and systemic barriers (e.g., syllabus constraints, educational inequities), aligning with Vygotsky's (1978) Zone of Proximal Development and Anderson's (2002) Curriculum Alignment Theory.

1. How do you perceive the alignment between Vision 3's grammar presentation (e.g., boldfaced extracts, rule charts, drills) and the rigor of national English exam questions? (Probe: Provide an example, such as passive voice.)
2. What challenges do students face when navigating the gap between Vision 3 grammar exercises and national exam demands? (Probe: How does this affect their exam preparedness?)
3. What metacognitive strategies do you teach to support students' grammar preparation for the exam? (Probe: How do you encourage planning or self-evaluation in your instruction?)
4. How do curriculum factors, such as time constraints or syllabus scope, influence your ability to teach grammar effectively? (Probe: How do holidays or Ministry policies impact alignment with exam demands?)
5. What systemic barriers, such as educational inequities in state schools, hinder students' performance on the national exam? (Probe: How do Ministry policies on grade progression affect Vision 3 instruction?)