Taiwan Journal of TESOL
Vol. 23.2, 83-134, 2026 DOI: 10.30397/TJTESOL.202610_23(2).0003

GRADED GRAMMAR SEQUENCING AND THE CEFR: A CORPUS-BASED FRAMEWORK FOR TEXTBOOK REVIEW

Tzu-Fen Nellie Yeh

ABSTRACT

This study profiles grammar sequencing in Taiwan's K-12 English textbook system using the CEFR-related English Grammar Profile (EGP) as an external descriptive reference. A corpus of 57 student textbooks (Grades 3–12) was compiled, and 244 grammar points (GPs) were identified and mapped onto 602 consolidated EGP descriptors (A1–B2). Descriptor-based indices were used to quantify coverage, cross-series variation, cross-stage recycling, complexity gain, and early or isolated introductions, analyzed via chi-square tests, ANOVAs, and logistic mixed-effects models. Results show stage-aligned progression (primary A1; junior high A2–B1; senior high B1–B2), but only 244/602 descriptors (≈40%) appear, and 182/354 B1–B2 descriptors (≈51%) are absent as explicit grammar input. Cross-series differences within each stage are substantial. Although 42 GPs recur across stages, only 9 (≈21%) show clear complexity gain, and mixed-effects modeling indicates recycling is concentrated in intermediate bands (especially B1). A few A2 forms appear early and a few B2 forms appear late and in isolation, but omissions are more pervasive. The findings show how CEFR-linked descriptors can support systematic auditing of grammar strands and inform an EGP-based scope–level–recycling checklist for CEFR-oriented textbook design and review.

Key words: English Grammar Profile, grammar sequencing, corpus-based textbook analysis, spiral curriculum

Tzu-Fen Nellie Yeh, Research Center for Translation, Compilation, and Language Education, National Academy for Educational Research, Taipei, Taiwan. (nellieyeh@gmail.com)

INTRODUCTION

Alignment with internationally recognized standards, often with the Common European Framework of Reference for Languages (CEFR) as a shared point of reference, has become a prominent policy goal in language education. In this study, the CEFR is understood as a descriptive reference framework, and national curricula, assessment systems, and textbooks are treated as policy instruments that selectively draw on its levels and descriptors in locally situated ways. Within CEFR-informed curricula, graded grammar syllabuses are frequently justified in terms of developmental sequencing, that is, introducing grammatical resources in step with learners' expanding communicative repertoires and cognitive capacities (DeKeyser, 2017; Ellis, 2006; Pienemann, 1998). At the same time, comprehension-based approaches associated with the input hypothesis emphasize that acquisition can proceed without detailed pre-planned grammar sequencing if learners receive abundant, meaningful, comprehensible input (Gass, 2015; Krashen, 1992, 1997, 2017; Lichtman & VanPatten, 2021). The present study takes these divergent perspectives seriously by treating textbook grammar sequencing as one controllable dimension of input design.

Recent work on CEFR-referenced reforms in Asia suggests that alignment has progressed further for vocabulary, reading, and assessment than for grammar, and that the translation of CEFR-related aims into classroom practice is mediated by national curricula, textbook policies, and teacher beliefs (Ahmad Afip et al., 2019; Fives & Buehl, 2016; Jeon, 2022; Savski, 2020). System-level studies of grammar progression in CEFR-referenced textbooks remain scarce, although emerging research in Malaysia, Vietnam, and Turkey has begun to document misalignment between intended proficiency bands and textbook input, late clustering of complex structures at the end of schooling, and limited recycling across grades (Azaha et al., 2024; Kulaklı, 2024; Nguyen, 2024; Sulaiman & Singh, 2021). For Taiwan, where CEFR-related curriculum reform is crystallizing in the development of a localized CEFR-T framework, there is still no comprehensive description of how grammar points (GPs) are staged across the full K-12 span in officially approved textbooks, despite

their central role in structuring classroom input.

Existing studies on CEFR implementation in Taiwan point to discrepancies between policy rhetoric and classroom realities, including constrained teacher agency in materials use, inconsistent CEFR labeling of textbooks, and uncertainty about how proficiency levels should inform everyday planning (Haney et al., 2002; Hong et al., 2020; Hsieh, 2022; Timperley & Alton-Lee, 2008). In a system where many teachers rely heavily on student books to organize instruction, particularly in large, exam-oriented classes, understanding the trajectories of grammar input encoded in those books is a necessary step toward assessing what kinds of learning opportunities CEFR-related curriculum reforms actually make available to school-age learners.

Responding to this gap, the present study offers a corpus-based analysis of GP distribution and progression in Taiwan's official English textbooks for primary, junior high, and senior high school. Rather than benchmarking series directly against the full A1–C2 scale, the analysis adopts a stage-sensitive mapping model with three overlapping bands: A1–A2 for primary, A2–B1 for junior high, and B1–B2 for senior high. These bands are used heuristically to characterize stage-appropriate coverage and expected developmental trajectories in light of national curriculum goals. Textbook-intended GPs are identified and consolidated through a multi-rater annotation protocol and then related to CEFR levels using the English Grammar Profile (EGP; O'Keeffe & Mark, 2017) as a corpus-informed descriptive resource for profiling textbook grammar input in Taiwan (Baron & Papageorgiou, 2014). This design makes it possible to examine not only which GPs are included and when they first appear, but also how far they are recycled with increased functional range and processing demands in line with usage-based and practice-oriented accounts of development (DeKeyser, 2017; Ellis, 2006; Spada & Tomita, 2010).

The contribution of the study is twofold. Theoretically, it clarifies how grammar progression is currently realized in Taiwan's nationally approved K-12 English textbook system developed under the 2019 Curriculum Guidelines, and uses CEFR-related descriptors as external reference points to build a stage-sensitive account of sequencing and

recycling that links these descriptors with second language acquisition (SLA)-informed criteria, including processing constraints, practice conditions, and input considerations. Practically, it provides system-level evidence to inform textbook developers, curriculum reviewers, and policymakers in aligning grammar strands with CEFR-related goals while maintaining a focus on learnability and learner development, without assuming that tighter CEFR alignment is always preferable. Accordingly, the study addresses the following research questions:

1. How are GPs associated with CEFR A1–B2 presented and distributed across primary, junior-high, and senior-high textbooks in Taiwan?
2. How do GP coverage, sequencing, and recycling vary across the three Ministry-approved textbook series at each schooling stage?
3. To what extent do cross-stage trajectories of GPs appear developmentally plausible when evaluated using criteria informed by Processability Theory, Skill Acquisition Theory, form-focused instruction, and input-oriented perspectives, together with CEFR-related level information?
4. In what ways do the textbook series exhibit misalignment with the stage-sensitive model, redundancy without substantial complexity gain, or gaps in GP coverage across stages?

Situated within usage-based SLA frameworks and informed by both sequencing-oriented and input-oriented views of acquisition, this study addresses the underexplored challenge of graded grammar progression in CEFR-referenced curricula and provides empirical evidence to inform debates about how international reference frameworks are realized in local textbook and syllabus design.

LITERATURE REVIEW

Research on CEFR-referenced curricula has increasingly examined how level descriptors can inform the design of graded grammar syllabuses in EFL settings, including through corpus-based

tools such as the EGP (Baron & Papageorgiou, 2014; O'Keeffe & Mark, 2017).

Building on this foundation, the review examines CEFR localization in curriculum and textbooks, the role of EGP-style descriptors in stage-based grammar profiling, empirical work on grammar progression, SLA perspectives on sequencing and practice, and remaining system-level gaps in CEFR-referenced materials, with a focus on Taiwan.

CEFR in Curriculum Development and Localization

As the CEFR has been adopted in an expanding range of education systems, its level descriptors have been used to orient curriculum goals, assessment scales, and textbook series well beyond Europe (Jeon, 2022; Koizumi, 2022; Nagai et al., 2020; Uri, 2021). In these reforms, the CEFR and its Companion Volume are framed as flexible mediating tools whose concrete consequences depend on how national and local actors interpret and operationalize their descriptors, rather than as fixed standards that directly dictate classroom practice (Little & Figueras, 2022; North & Piccardo, 2023).

Comparative reviews of CEFR-aligned reforms in the ASEAN region similarly emphasize that the framework shapes classrooms only to the extent that it is embedded in coherent curricula, assessment systems, and professional learning structures (Hamid et al., 2025; Ng et al., 2022).

In non-European contexts, alignment of grammar instruction with CEFR levels and the translation of policy into everyday teaching have proved uneven, and linking graded grammar syllabuses to CEFR levels has emerged as a more complex and less fully developed area than the better-documented alignment of vocabulary and reading components (Jeon, 2022; Savski, 2020). Malaysia's English Language Education Reform Roadmap 2015–2025 illustrates these challenges. Ahmad Afip, Hamid, and Renshaw (2019) report partial alignment for reading and listening but ongoing difficulties in speaking and writing, associated with teacher preparedness and limited localized materials; reliance on imported CEFR-labeled textbooks widened the gap between policy and classroom realities.

Across Southeast Asia, Savski (2020) argues that global reference frameworks only shape classroom practice insofar as they are mediated through context-sensitive processes such as stakeholder consultation, curriculum coherence, and sustained teacher development, a conclusion echoed in broader critiques of validation and implementation practices in language testing and curriculum reform (Im et al., 2019). In South Korea, despite CEFR-inspired assessment reform and professional development, Jeon (2022) documents persistent policy–practice gaps, with many teachers feeling underprepared to work with CEFR levels in everyday instruction.

Japan's experience with the localized CEFR-J further highlights the grammar issue. CEFR-J has supported more fine-grained assessment practices and the development of leveled wordlists and functional expressions (Negishi et al., 2013; Tono & Negishi, 2020), yet its elaboration has been more advanced for vocabulary and formulaic expressions than for grammar, and a comparably detailed, empirically grounded grammar strand remains underdeveloped (Shao, 2019; see also Nagai et al., 2020, for broader CEFR-informed applications). These cases underscore that CEFR-related resources serve as descriptive points of reference that must be interpreted, adapted, and supplemented. This is particularly true for grammar, which depends on locally developed curricula, materials, and teacher support mechanisms.

In Taiwan, CEFR-related reforms have coalesced around CEFR-T, a localized framework intended to integrate CEFR principles into the national K-12 English curriculum. Recent analyses of English education under globalization point to tensions between policy aims, curriculum design, and classroom realities, including the status of English as a lingua franca and the pressure of high-stakes examinations (Ke, 2021; Lin, 2020). Empirical research on GP distribution and progression in official textbooks, however, remains limited. Hsieh (2022) identifies discrepancies between policy discourse and classroom practice, including constrained teacher agency, misalignment in textbooks, and uncertainty about proficiency levels. Hong, Tseng, Peng, and Sung (2020) report inconsistencies in textbook CEFR labeling, pointing to the need for clearer and more

reliable criteria for alignment, while corpus-based evaluations of senior-high English textbooks suggest that text difficulty does not always match the demands of college entrance examinations (Cheng & Chang, 2022).

These system-level patterns intersect with a growing body of Taiwan-based research on teacher beliefs and classroom enactment. Studies of teachers' orientations toward learner autonomy, writing instruction, and alternative curricula consistently show that classroom practices are filtered through local beliefs, institutional pressures, and exam expectations (Chang, 2020; Chen, 2018; Chiang, 2003; Chien, 2023; Lin, 2024; Tseng, 2019). Against this backdrop, there is a need to examine how Taiwanese textbooks interpret and operationalize CEFR-related curriculum goals when sequencing grammar within the evolving CEFR-T policy environment.

The EGP and Stage-Based Grammar Descriptors

The EGP offers empirically derived descriptors of grammar use calibrated to CEFR levels A1–C2 and has become a central resource for characterizing how particular forms and functions tend to appear across the proficiency scale (Baron & Papageorgiou, 2014; O'Keeffe & Mark, 2017). Developed by Cambridge University Press and Cambridge English Language Assessment, it draws on large learner corpora to identify recurrent patterns in authentic learner production, providing a corpus-based view of how grammatical resources are distributed over levels. In the present study, EGP categories are used as corpus-informed, level-related reference points for profiling textbook grammar input, understood as an initial approximation to developmental staging that must be interpreted in light of local learner data.

A distinctive feature of the EGP is its meaning-oriented organization. Instead of grouping items as decontextualized forms, descriptors are phrased as "can do" statements that link form to communicative function at each level (O'Keeffe & Mark, 2017). For example, at B1 learners are described as using modal verbs to express obligation, while at B2 they extend modal constructions to express probability or hypothetical stance. This functional trajectory resonates

with the CEFR's action-oriented orientation and with recent elaborations of illustrative descriptors in the Companion Volume (Little & Figueras, 2022), and is particularly useful for evaluating whether textbook sequences support a gradual expansion of learners' options for expressing stance, time, and textual relations. Parallel developments in CEFR-J-related resources, such as the grammar- and function-oriented materials compiled for Japanese contexts (Negishi et al., 2013; Tono & Negishi, 2020), likewise show how leveled descriptors can be organized around meaning rather than isolated forms.

Building on this orientation, EGP-style profiling has been used to inspect how grammar input is distributed across levels in textbooks and curricula, and to identify cases where local materials over- or under-represent structures relative to targeted bands (Aman, 2021; Sulaiman & Singh, 2021). Similar descriptor-based approaches have supported the mapping of national standards onto the CEFR (Little & Figueras, 2022; North & Piccardo, 2023) and the design of CEFR-informed tests and tasks (Nagai et al., 2020). Parallel work with CEFR-J resources illustrates comparable possibilities on the lexical side: the Vocabulary Analyzer Based on CEFR-J Wordlist for Self-Reflection (VACSR) Version 2 uses CEFR-J-based wordlists to profile learner production and provide level-related feedback, while also showing that level assignments are conditioned by the underlying classroom corpora and tasks (Ohashi et al., 2023). Together, these applications demonstrate that external descriptor systems can support curriculum auditing and alignment, but only when their level information is interpreted with regard to local contexts and learner profiles.

At the same time, corpus-based descriptor systems such as the EGP do not automatically translate into pedagogical practice or classroom assessment. Valax (2011) emphasizes that teachers need support to interpret leveled descriptors and to use them in planning sequences of activities, and research on validation and classroom-based assessment likewise cautions that frameworks and profiles must be mediated through local assessment cultures, teacher cognition, and institutional constraints if they are to inform real decisions about teaching and testing (Alsalem, 2020; Im et al., 2019; Lovrić, 2024).

Recent discussions of CEFR- and CEFR-J-related work echo this view, stressing that level labels require contextual mediation if they are to inform day-to-day teaching rather than simply rebrand existing practices (Little & Figueras, 2022; Nagai et al., 2020). When incorporated into teacher education and local materials development, EGP-style descriptors can therefore function as a descriptive bridge between CEFR-related benchmarks and classroom grammar goals, helping to articulate syllabuses that are sensitive to level and oriented toward learner development. In the present study, this bridging potential is taken up more fully in the analytic framework outlined in Methodology.

Empirical Studies on Grammar Progression in Textbooks

Within this broader body of CEFR-referenced work, a growing subset of studies examines how grammar is selected, graded, and sequenced in ESL/EFL textbooks. Corpus-based analyses suggest that trajectories of grammatical complexity do not always keep pace with other dimensions of input. Lynn (2021), for instance, found that in a corpus of ESL/EFL reading textbooks, increases in lexical sophistication were not accompanied by comparable growth in grammatical complexity, limiting systematic support for the expansion of learners' grammatical repertoires. Burton (2020), comparing textbook progressions with SLA findings, similarly argues that many conventional grammatical syllabuses diverge from observed developmental paths and calls for sequencing that more closely reflects empirically attested acquisition patterns.

In Asian curriculum contexts, comparative evaluations have highlighted problems of range and progression in grammar coverage. Kulaklı (2024) reports that Turkey's Grade 12 curriculum and coursebook concentrate substantial numbers of advanced GPs at the end of schooling without sufficient preparatory coverage at earlier stages, undermining coherence across levels. Nguyen (2024) shows that Vietnamese secondary textbooks often lack an explicit rationale for grammar progression, with sequencing appearing to follow editorial or thematic considerations more than a principled view of cumulative learning. Systematic reviews of CEFR-aligned curricula

in the ASEAN region similarly point to uneven integration of descriptors into actual syllabuses and materials, particularly for grammar strands (Hamid et al., 2025; Ng et al., 2022). Together, these studies indicate that, where developmentally informed reference points are not used to guide selection and ordering, textbook grammar input may fail to provide learners with a coherent pathway for extending their grammatical and communicative resources over time.

Teacher-focused research reinforces this picture by documenting how teachers respond when textbook sequencing is perceived as misaligned with learner needs. Chali (2020) found that Ethiopian EFL teachers, while strongly valuing grammar instruction, frequently departed from the order presented in textbooks when they judged it inappropriate for students' current abilities. In flipped classrooms in Saudi Arabia, Albahouth (2024) reports that instructors routinely reorder, supplement, or skip units in order to sustain engagement and make grammar work more meaningful. Studies of teacher cognition in grammar assessment similarly show that teachers negotiate between formal expectations and their own judgments about processing requirements when deciding what to test and when (Alsalem, 2020; Lovrić, 2024). Parallel work in Taiwan, discussed in CEFR in Curriculum Development and Localization, likewise portrays teachers as active mediators of curriculum and materials, whose beliefs about autonomy, writing instruction, and language policy shape how textbook content is re-sequenced and adapted in practice.

Despite these contributions, large-scale analyses that track grammar input across a full K-12 span remain rare. Most textbook studies focus on individual grades, specific skills, or a single series, and only a small subset explicitly relates grammar coverage to CEFR levels. CEFR alignment has been more extensively examined in assessment domains than in the design of grammar strands in core teaching materials, particularly through studies of CEFR-referenced reading tests and text difficulty (Cheng & Chang, 2022; Kareema, 2023; North & Piccardo, 2023; Vettori et al., 2024). As a result, descriptions of how grammar syllabuses unfold across grades and series remain fragmentary, and Taiwan is no exception: existing analyses of curricula and textbooks highlight tensions around

globalization, English as a lingua franca, and exam preparation (Cheng & Chang, 2022; Ke, 2021; Lin, 2020), but do not yet map graded grammar progression across stages or series in CEFR-related terms.

SLA Perspectives on Grammar Sequencing

Principles for grammar instruction in EFL contexts are often grounded in Processability Theory (PT), Skill Acquisition Theory (SAT), and research on form-focused instruction (FFI). These perspectives converge on the view that acquisition proceeds through partly predictable sequences and that instruction is most effective when it is organized to create sustained opportunities for learners to work with target forms in meaningful use (DeKeyser, 2017; Pienemann, 1998; Spada & Tomita, 2010).

PT (Pienemann, 1998) proposes that learners acquire grammatical structures in an order constrained by their current processing procedures: operations at the lexical and phrasal level are acquired before procedures that support subject–verb agreement, subordination, or clause embedding. Empirical work with typologically diverse learner groups, including Chinese L2 learners (e.g., Wang, 2011), has documented hierarchies in which attempts to teach structures that require higher-level procedures before simpler patterns have stabilized rarely lead to durable gains. For curriculum and textbook design, this suggests that sequences should avoid large jumps in processing demands and that new forms are most productively introduced when learners have had opportunities to consolidate prerequisite patterns. In the present study, PT is therefore used as a heuristic for judging whether textbook progressions respect plausible processing constraints, rather than as a strict template against which stages are formally tested.

FFI research highlights the timing and integration of form and meaning. Spada and Tomita's (2010) meta-analysis shows that instructional effects depend not only on which features are targeted but also on how instruction creates opportunities for development through input enhancement, feedback, and tasks that require learners to deploy target forms to achieve communicative goals (see also Gass,

2015). Studies comparing explicit and more implicit approaches suggest that direct attention to form can facilitate learning when it is embedded in meaningful activity, but that sequencing based solely on formal difficulty is insufficient if learners are not positioned to use the forms for real communicative purposes (Freedman, 1993; Grahn Sherin, 2023).

SAT (DeKeyser, 2017) accounts for how learners move from declarative knowledge about forms to proceduralized, automatized use through practice. It distinguishes highly controlled practice, structured communicative activities, and more open-ended tasks, and predicts that gains in fluency and accuracy arise when practice gradually shifts from low-variability, form-focused work to higher-variability, meaning-focused use. Applying SAT to activity design, Criado (2016) argues for sequences of tasks that systematically revisit target forms while increasing processing demands. This view underpins the notion of complexity gain used later in this paper, where revisiting a GP is considered developmentally valuable only when it is associated with expanded functional range, tighter discourse integration, or higher processing load, rather than simple repetition of earlier uses.

At the same time, not all SLA perspectives assign a central role to pre-planned grammar sequencing. Comprehension-based approaches, particularly those associated with Krashen's input hypothesis, emphasize that acquisition is driven primarily by exposure to meaningful, comprehensible input rather than by explicit syllabuses of forms (Krashen, 1992, 1997, 2017). Subsequent work has refined and critiqued this position, debating how far explicit instruction and output opportunities contribute beyond input alone (e.g., Gass, 2015; Jegerski, 2021; Lichtman & VanPatten, 2021; Loschky, 1994; White, 1987). Reviews of explicit versus implicit grammar instruction likewise suggest that explicit focus on form can be beneficial under some conditions but is neither necessary nor sufficient by itself (Freedman, 1993; Grahn Sherin, 2023). Against this backdrop, the present study does not assume that carefully graded grammar syllabuses are inherently superior to input-rich approaches; instead, textbook sequencing is treated as one controllable dimension of input design whose developmental value depends on the opportunities it

affords for form–meaning use.

Taken together, these SLA perspectives support viewing grammar progression as staged, usage-based, and practice-rich, while recognizing that acquisition also depends on the quantity and quality of comprehensible input available to learners. In this study, PT, SAT, FFI, and input-oriented perspectives jointly provide a descriptive lens for evaluating whether the grammar sequences and recycling patterns in Taiwanese textbooks offer coherent, realistically staged trajectories of development, without presupposing that CEFR-aligned sequencing is either a necessary or a sufficient condition for successful learning.

Research Gaps and Rationale for the Present Study

Across Asian systems that have adopted CEFR-referenced curricula, empirical work has only selectively examined how grammar is staged in mainstream materials. Studies have tended to foreground vocabulary lists, reading difficulty, or test specifications (Azaha et al., 2024; Lynn, 2021; Uri, 2021), reflecting a broader tendency for CEFR implementation and research to focus on reading and assessment domains rather than on graded grammar strands (Hamid et al., 2025; Little & Figueras, 2022; Nagai et al., 2020; North & Piccardo, 2023; Vettori et al., 2024). Much less attention has been paid to how GPs are selected, distributed, and linked to proficiency bands over the course of schooling. Even where CEFR is formally invoked, grammar strands often appear thinly scaffolded in comparison with other domains, exhibiting overload of structurally demanding forms, out-of-sequence introductions, and weak cumulative progression (Aman, 2021; Jeon, 2022; Kulaklı, 2024; Nguyen, 2024; Sulaiman & Singh, 2021). Nguyen (2024) further suggests that some progressions seem to reflect editorial convenience rather than an explicit theory of gradual, cumulative learning. These patterns underscore the lack of system-level descriptions of grammar input trajectories in CEFR-referenced materials, a concern echoed in reviews of CEFR-informed curricula and assessment practices that point to gaps between stated level descriptors and the actual design of teaching materials (Hamid et al., 2025; Im et al., 2019; Ng et al., 2022).

This makes it difficult to judge whether learners encounter a

coherent grammar pathway from beginner through intermediate levels, or whether early learning is systematically recycled and extended. Sulaiman and Singh (2021) therefore argue that effective grammar profiling must address both the selection of target structures and their pedagogical sequencing over time; without a K-12 perspective, gaps, redundancies, and abrupt jumps in demands remain largely invisible. At the same time, some SLA perspectives emphasize that rich, comprehensible input may matter more for acquisition than detailed pre-planning of grammatical sequences (Bell & Payant, 2024; Krashen, 1992, 1997, 2017). From this vantage point, what is missing is not another defense of tightly scripted grammar syllabuses, but empirically grounded accounts of the kinds of grammar input trajectories that current CEFR-referenced textbooks in fact offer learners across a full K-12 span, and how these trajectories interact with existing assessment regimes and classroom practices (Im et al., 2019; Lovrić, 2024).

In Taiwan, this gap is particularly marked. Despite ongoing CEFR-T development and successive textbook reforms, research has largely concentrated on curriculum intentions, English-as-a-lingua-franca orientations, and exam-related text difficulty (Cheng & Chang, 2022; Ke, 2021; Lin, 2020), alongside studies of teacher beliefs and classroom mediation of policy (see CEFR in Curriculum Development and Localization and Empirical Studies on Grammar Progression in Textbooks). There has been no system-level analysis of how grammar input is staged across primary, junior-high, and senior-high textbooks in relation to CEFR-related descriptors and SLA-informed criteria, a gap the present study directly addresses. To do so, the study systematically maps GPs in Taiwan's official English textbooks across the twelve years of compulsory education, using a stage-sensitive model to examine their frequency, sequencing, and the extent to which observed patterns appear likely to support cumulative development for school-age learners. By integrating EGP descriptors with PT-, SAT-, and FFI-informed views of developmental trajectories, the study develops a CEFR-referenced analytic framework that profiles how grammar is currently graded in policy textbooks, using CEFR-related resources as descriptive tools rather than prescriptive targets that materials or learners are expected to match (Little &

Figueras, 2022; Nagai et al., 2020; North & Piccardo, 2023). While centered on Taiwan, this framework is designed to be portable: misalignment issues reported in Malaysia and Turkey (Azaha et al., 2024; Kulaklı, 2024) suggest that the same methodology could be used to audit grammar grading across schooling stages in other CEFR-referenced systems and to support cross-national comparisons of textbook-based grammar input.

METHODOLOGY

Research Design

A corpus-based, cross-stage comparative design was adopted to examine how GPs are selected, ordered, and recycled in Taiwan's K-12 English student books. Student textbooks were treated as a major, policy-mediated source of input within a one curriculum, multiple textbooks system, and developmental plausibility was evaluated primarily against SLA criteria, with CEFR-related information used as a descriptive reference rather than as a prescriptive standard. In this framework, CEFR labels function as corpus-informed indices of the nominal level range of textbook input, whereas developmental appropriateness is judged in relation to SLA perspectives on processing demands, practice conditions, and opportunities for form–meaning use.

Three schooling stages were distinguished in accordance with the structure of Taiwan's national curriculum and the associated textbook approval process: primary (Grades 3–6), junior high (Grades 7–9), and senior high (Grades 10–12). Policy documents linked to CEFR-T position primary, junior high, and senior high schooling with cumulative attainment around A1, A2, and B1, respectively. Building on the CEFR Companion Volume, which conceptualizes levels as overlapping bands rather than hard cut-offs (Council of Europe, 2020), and on Vygotskian views of scaffolded progression and the zone of proximal development (ZPD; Vygotsky, 1978), the analysis treated primary as targeting the A1–A2 band, junior high A2–B1, and senior high B1–B2. These bands are interpreted as flexible target zones in which learners predominantly work with forms typical of that range

while also encountering some structures above and below it.

For each stage, all grammar-designated sections in student books from the three Ministry-approved series available at that level were digitized and compiled into stage-specific corpora. Within these corpora, GPs were defined as textbook-intended grammatical targets, identified on the basis of section labels, exemplar sentences, and accompanying practice activities, and then consolidated through a multi-rater coding procedure. Each consolidated GP was linked to a CEFR level using the EGP as an external, corpus-based descriptor system; EGP levels were treated as heuristic labels for typical learner usage patterns rather than as benchmarks that textbooks or learners are expected to match.

To avoid circularity between CEFR-based mapping and evaluation, and to separate scope from sequencing, the analysis proceeded along three dimensions. First, coverage was operationalized as the set of distinct GPs associated with CEFR A1–B2 that appeared in each stage and series. Second, sequencing was defined as the order and grade level of first introduction for each GP, allowing examination of whether new forms constituted sizable processing jumps or extended established patterns in ways broadly consistent with PT-, SAT-, and FFI-informed expectations, treated here as qualitative benchmarks rather than formal hypotheses. Third, recycling was examined by tracking whether GPs reappeared across grades and stages with increased functional range, tighter discourse integration, or higher processing load. Such developments were coded as complexity gain, in contrast to simple repetition. Together, these three strands yield a stage-sensitive index of how Taiwanese textbooks structure grammar input, which is later interpreted with reference to SLA-based criteria and locally interpreted CEFR-related curriculum goals.

Corpus and Materials

The corpus comprised 57 English student textbooks covering Grades 3–12 in Taiwan. All volumes were developed in line with the 2019 national English curriculum drafted by the National Academy for Educational Research and approved by the Ministry of Education,

and had passed a national review process based on transparent, criteria-referenced evaluation of content quality. For each schooling stage, three Ministry-approved series available at that level were included: P-KH (Kang-Hsuan), P-HL (Han-Lin), and P-HE (Hess) for primary (Grades 3–6); J-KH (Kang-Hsuan), J-NY (Nan-Yi), and J-HL (Han-Lin) for junior high (Grades 7–9); and S-SM (San-Min), S-LT (Long-Teng), and S-HL (Han-Lin) for senior high (Grades 10–12). Each primary series comprised eight semester-based volumes, each junior-high series six volumes, and each senior-high series five volumes, yielding 24, 18, and 15 volumes, respectively (57 in total). At each stage, these three series constituted the complete set of Ministry-approved options available to schools during the period of data collection, and schools typically adopted one series per stage for the duration of that stage. In the statistical analyses, series refers to parallel textbook lines within each stage, rather than to a single publisher's continuous K-12 coursebook line, reflecting the way textbooks are selected and adopted under Taiwan's current curriculum policy.

The analysis was deliberately restricted to student books rather than to workbooks, teacher guides, or digital supplements. In Taiwan's approved textbook system, the student volume constitutes the common, compulsory written input that all enrolled learners are expected to encounter, whereas ancillary components vary substantially across publishers, schools, and teachers. Workbooks typically present fragmentary sentences and discrete-item exercises oriented toward practice rather than stable input; teacher manuals and digital materials contain publisher-specific extensions that teachers may or may not adopt. Focusing on student books therefore provides a conservative and comparable view of the core grammar input encoded in policy-sanctioned materials, while recognizing that additional grammar work may occur elsewhere. This limitation is revisited in the Discussion.

Within each volume, only sections in which grammar was an explicit curricular focus were annotated. These typically appeared under headings such as "Grammar Focus," "Sentence Patterns," "Language Use," or "Grammar Review"; at the primary level, introductory units and Classroom English pages were also included

when they introduced systematic patterns rather than isolated phrases. At the junior- and senior-high levels, grammar sections were usually positioned after reading texts and in unit or end-of-book review pages. Concentrating on these dedicated slots aligned the corpus with the textbook-intended grammar syllabus and avoided conflating it with incidental language in dialogues or reading passages.

To define the potential grammar space in a principled way, a reduced pool of EGP descriptors covering CEFR levels A1–B2 was first constructed. The reference inventory of level-related descriptors was taken from the online EGP database, developed within the Cambridge English Profile project, and included all A1–B2 entries available in the public database at the time of data collection. Because the EGP is periodically updated, all analyses reported here are based on a static snapshot of the A1–B2 entries downloaded in September 2024; later revisions of the online database are not reflected in the coding. The original EGP inventory for A1–B2 comprised 981 descriptors (A1 = 109, A2 = 291, B1 = 338, B2 = 243). The consolidation procedure applied in this study reduced this set to 602 pedagogically distinct units, comprising 73 at A1, 175 at A2, 201 at B1, and 153 at B2, while retaining level-related contrasts where they captured meaningful differences in form–function mappings. Of these 602 units, 244 were attested at least once in the 57 textbooks analyzed; the remaining potential descriptors did not occur in the student books and therefore do not enter into later coverage or sequencing indices.

Consolidation was guided by a set of a priori rules designed to retain pedagogically meaningful distinctions while removing redundancy. Exact duplicates and non-substantive entries were eliminated; highly overlapping or fully subsumed descriptors were merged; items that differed only in polarity or interrogative/affirmative form but expressed the same function (e.g., *can* for ability) were collapsed; and descriptors that solely reflected lexical expansion without a distinct grammatical pattern were excluded. Pattern-only labels (e.g., "X *than* Y" without a clear functional specification) were linked to functionally richer descriptors when they referred to the same underlying construction, and broad umbrella entries were dropped when more specific, level-calibrated descriptors existed. A worked example at A1 illustrates this

consolidation logic. Five closely related descriptors were merged into a single teachable unit. These were A1-53 "Can use the affirmative question form of *can,*" A1-54 "Can use *can* to talk about ability to do something," A1-55 "Can use *can* to make an offer," A1-56 "Can use *can* to talk about possibility," and A1-57 "Can use *can* to make a request." The resulting unit was A1-36 "Can use the interrogative form of *can* to express basic functions such as ability, offers, requests and possibility." These uses share the same interrogative pattern and are typically taught as a single lesson in Taiwanese textbooks, whereas modal meanings such as permission follow a different structural frame and were therefore kept separate (see Annotation Procedure).

Within this descriptor space, a GP was operationalized as a textbook-intended target structure that could be mapped onto at least one consolidated EGP descriptor. When scanning each grammar section, coders extracted all such targets, cross-checked them against the 602-unit pool, and grouped textbook realizations that instantiated the same descriptor as a single GP. Across the 57 volumes, this procedure yielded 244 distinct GPs, corresponding to the subset of the 602 consolidated EGP units that were attested at least once in the textbooks. These 244 GPs constitute the units of analysis for the subsequent coverage, sequencing, and recycling measures. The consolidation rules and category definitions were recorded in the coding manual used by annotators during the procedures described in Annotation Procedure.

Annotation Procedure

Annotation proceeded in three independent rounds conducted by three coder teams, each working across all three schooling stages (primary, junior high, senior high). In this way, every grammar-designated section received at least three independent readings before consolidation, and all items on which coders disagreed were carried forward to a final reconciliation phase.

All coders participated in a preparatory workshop prior to data collection. The training introduced the consolidated EGP descriptor set and the operational definition of a GP as a textbook-intended target structure anchored in explicit labeling, exemplifying sentences, and

practice activities. Using textbook pages not included in the corpus, coders practiced identifying GPs and assigning EGP levels, focusing on borderline phenomena such as different uses of the present perfect, layered modal meanings, and overlapping clause types. These discussions informed a written coding manual specifying decision rules for when clusters of related forms (e.g., multiple uses of would in one lesson) should be treated as a single GP or split into separate GPs, and how to select an EGP level when a descriptor family spanned more than one band.

During the main annotation phase, coders worked independently from this manual. For each candidate GP, they recorded the textbook stage, series, and unit, at least one attested example sentence, and the corresponding EGP descriptor code and CEFR level. When a textbook treatment combined structurally similar items but clearly distinct functions or meanings (e.g., *can* for ability vs. permission), separate GPs and level mappings were assigned; conversely, formally different variants realizing the same function (e.g., affirmative/ negative/ interrogative patterns within one descriptor) were coded as a single GP.

Particular attention was paid to features that span multiple EGP levels, such as the present perfect. In such cases, each instance was judged in its immediate instructional context (task and co-text) so that the assigned level reflected function-in-use rather than form alone. Present perfect forms used to convey decontextualized life experiences (e.g., *She has been to Japan once*) were mapped to lower-band descriptors, whereas uses integrating temporal framing, contrasts with the past simple, or narrative sequencing (e.g., *She has lived in Tokyo for ten years but moved last year*) were mapped to higher-band descriptors that explicitly capture these discourse functions. When contextual cues clearly supported a higher-level reading, that higher level was assigned; when the reading remained ambiguous, the lower, structurally defined level was used to avoid inflating developmental claims. If a sentence simultaneously instantiated more than one distinct feature (e.g., an A1 preposition plus an A2 progressive aspect), it received all applicable tags, but where a single construction allowed competing interpretations for the same function, only one level was retained following the same

precedence rule.

After independent coding was completed, the three teams met for a series of adjudication sessions. All discrepancies in GP identification or CEFR assignment were systematically reviewed, drawing on the consolidated EGP inventory, example sentences, and the coding manual. In cases where coders had assigned adjacent levels (e.g., A2 vs B1) to the same structure, the team revisited both the textbook context and the wording of the relevant EGP descriptors before agreeing on a single best-fit level. Where necessary, the manual was refined to record new decision rules so that similar cases would be treated consistently across the corpus. Inter-rater agreement on a stratified subset of the data was quantified using Cohen's κ prior to consensus.

This multi-stage procedure yielded a consensus-based dataset in which each GP is linked to its source textbook and stage, a representative attested example, and a confirmed CEFR level via an explicit EGP descriptor code. The resulting annotation provides a transparent bridge between local textbook syllabuses and externally derived level descriptors, treating these mappings as analytically motivated approximations rather than as claims about learners' actual mastery, and constraining subjectivity through a shared manual, explicit precedence rules, and multi-rater coding with quantified agreement.

Reliability and Validity Measures

Reliability and validity were examined at multiple points in the annotation process. Inter-rater consistency was first assessed on the independently coded dataset, prior to any reconciliation, and then complemented by an expert audit of a stratified sample of items. These checks covered GP identification, CEFR level assignment via EGP descriptors, and the additional coding of recycling and complexity gain.

For CEFR-level assignment, raw agreement across annotators on pre-consensus coding exceeded 90%, and Cohen's κ was .89 (95% CI [0.85, 0.93]). Agreement on GP identification and on recycling/ complexity-gain codes was likewise quantified on a stratified subset

of recycled items, yielding κ = .87. Across these coding decisions, κ estimates fell within what Landis and Koch (1977) term the substantial to almost perfect range, indicating that judgments about what counted as a distinct GP, whether it was recycled, and whether later occurrences represented genuine complexity gain were not idiosyncratic to individual coders. All κ values were calculated on blind, independent pre-consensus coding and should therefore be interpreted as conservative, lower-bound estimates of reliability; subsequent consensus meetings further increased consistency but were not re-analyzed statistically.

To examine content validity, two external specialists in English language pedagogy and CEFR-based curriculum alignment reviewed a random 10% sample of the annotated entries. Using a structured checklist, they judged whether each segment had been correctly identified as a GP and whether the selected EGP descriptor and CEFR level were appropriate and sufficiently supported by the textbook context. The external audit showed that the large majority of mappings were endorsed as satisfactory, with auditors agreeing with approximately 88% of the GP–descriptor–level mappings; only a small minority of codes were flagged for revision, typically involving distinctions between neighboring descriptors or overlapping clause types. These contested cases were used to refine decision rules in the coding manual, in line with principles for CEFR-related content validation in textbook and test specifications.

Construct validity was considered in relation to both CEFR-based expectations and SLA-informed views of developmental staging. At an aggregate level, the annotated distributions met basic plausibility conditions: core A1 structures (e.g., simple present, simple past, high-frequency prepositions) predominated in primary textbooks, whereas more advanced forms (e.g., mixed and third conditionals, dense subordination) were concentrated in senior high materials and rarely appeared at primary. Patterns of recycling and complexity gain likewise aligned with the expectation that revisiting a GP at later stages should be associated with increased functional range or processing demands rather than simple repetition. At the same time, EGP-based level labels are derived primarily from non-Taiwanese learner corpora and have not been independently calibrated against

Taiwanese learners' performance. The CEFR levels used here are therefore interpreted as analytically motivated approximations to support systematic profiling of graded input in policy textbooks, rather than as definitive statements about actual difficulty or mastery for Taiwanese learners.

Data Analysis

The coded corpus was analyzed in two stages: first, by constructing descriptive profiles of grammar input by educational stage and textbook series, and second, by modeling stage- and series-related differences while accounting for the nested structure of the data.

At the item level, the annotations were reshaped into a matrix indicating, for each GP, the stage and series in which it first appeared as an explicit instructional target and whether it was subsequently revisited at later stages, with or without coded complexity gain. This structure distinguished coverage from sequencing. Coverage referred to whether a GP from the relevant EGP band appeared at least once in a given stage or series, whereas sequencing traced each GP's trajectory across stages, including the presence or absence of recycling and, for recycled items, whether later treatments met the predefined criteria for complexity gain.

Four sets of indices were derived from this matrix:

1. Stage-wise GP distribution and coverage. For each stage and series, the number and proportion of GPs mapped to A1, A2, B1, and B2 were calculated, together with coverage ratios defined as the proportion of EGP descriptors in the target band (e.g., A1–A2 for primary) represented by at least one textbook GP.
2. Recycling and complexity gain. A GP was coded as recycled if it was targeted as new content in more than one stage. For each recycled GP, later occurrences were classified as showing complexity gain or horizontal repetition, depending on whether subsequent treatment extended functional range, discourse integration, or processing demands beyond the

initial introduction. Frequencies and proportions of recycled items with and without complexity gain were computed by stage, series, and CEFR band.

3. Early and isolated introductions. Early introduction denoted GPs whose first occurrence fell below their mapped band (e.g., A2 GPs in primary, B1 GPs in junior high), whereas isolated introduction referred to B2 GPs appearing only in senior high without earlier staging. Absolute counts and percentages relative to all GPs in the corresponding band were calculated to gauge the magnitude of potentially premature or unscaffolded introductions.
4. Cross-series consistency. For each stage and CEFR band, the proportions of GPs shared by all three series, shared by exactly two series, or unique to a single series were computed, operationalizing equity-related concerns in a multi-publisher system by indicating the extent to which a common curriculum yields comparable grammar trajectories across series.

Subsequent statistical analyses treated coverage indices (e.g., proportions of EGP-derived units represented at each stage) separately from sequencing indices (e.g., timing of first introduction, prevalence of recycling with or without complexity gain), in order to disentangle breadth of exposure from the way that exposure was distributed over time. Table 1 summarizes how SLA- and CEFR-related work was translated into these evaluative dimensions.

Table 1

Criteria for Evaluating Grammar Sequencing (Derived from Prior Research)

Criterion	Supporting literature and rationale
Stage-appropriate sequencing	Grammar input should not substantially exceed learners' available processing procedures. Processability-based accounts describe a progression from lexical and phrasal operations to structures requiring agreement, subordination and clause embedding; large jumps in these processing demands are therefore treated as potential sequencing problems (Ellis, 2006; Pienemann, 1998).
CEFR-related coverage of core structures	CEFR descriptors and EGP profiles are used as corpus-informed reference lists for level-related GPs, indicating which key structures are present, clustered or absent at each stage, rather than as prescriptive targets that textbooks or learners must match exactly (Baron & Papageorgiou, 2014; Council of Europe, 2020; O'Keeffe & Mark, 2017).
Recycling and reinforcement of GPs	SAT predicts that movement from declarative to procedural knowledge depends on repeated, increasingly meaningful practice (DeKeyser, 2017). High-utility GPs are therefore expected to reoccur across units and stages, with spaced, varied opportunities for form–meaning use supporting consolidation (Criado, 2016; Spada & Tomita, 2010).
Vertical progression and complexity gain	When a GP is recycled at higher stages, progression is indicated by expanded functional range or greater syntactic/discourse complexity rather than simple restatement (Ellis, 2006; Pienemann, 1998). In this study, building on SAT, complexity gain is operationalized as such qualitative enrichment, whereas unchanged treatments are coded as horizontal redundancy (DeKeyser, 2017).

Criterion	Supporting literature and rationale
Cross-series consistency in a shared curriculum	In a single national curriculum with multiple approved series, broadly comparable access to core GPs is an equity concern (Savski, 2020; Tsui, 2020). CEFR-T documentation and EGP-based profiling function as common reference points that can support convergence on major level-related structures while allowing some pedagogical variation across publishers (Hong et al., 2020; Hsieh, 2022).

In addition to ANOVAs and contingency analyses on aggregated indices, a set of generalized linear mixed models (GLMMs) was fitted to examine cross-stage recycling patterns while explicitly modeling the hierarchical structure of the data. For these models, each GP was coded for whether it was recycled across stages (0 = taught as a target item in one stage only; 1 = explicitly retaught as a target item in a later stage). A logistic link was used, with CEFR level (A1, A2, B1, and B2) and textbook series entered as fixed effects, and stage specified as a random intercept to account for clustering by educational level. Wald χ^2 tests and odds ratios were used for the fixed effects, with statistical significance set at .05; model diagnostics indicated no major deviations from assumptions, minimal overdispersion, and no high-leverage outliers. Interpretation focused on converging patterns across analytic approaches and on effect sizes rather than on isolated *p*-values.

To provide more familiar and intuitive summaries, the GLMM findings were complemented by classical tests on aggregated indices. One-way ANOVAs compared mean counts of selected categories (e.g., A1 and B2 items, recycled-with-gain items) across stages and, where relevant, across series; assumptions were checked and, where violated, Welch corrections were applied. Effect sizes (η^2 or ω^2) and Bonferroni-adjusted post hoc tests are reported in the Results to qualify the practical significance of observed differences. A chi-square test of independence was applied to the contingency table of CEFR level by stage; where sparse cells threatened its validity, categories were collapsed or Fisher's exact test or log-linear models were used as robustness checks, with Cramér's *V* reported as an effect-size measure. All analyses were conducted in R 4.1 and SPSS

v26, with interpretation focusing on patterned effect sizes and converging evidence across analytic approaches to characterize the magnitude and distribution of alignment, misalignment, redundancy, and complexity gain across stages and textbook series.

RESULTS

Corpus Profile and Annotation Reliability

Across the 57 textbooks, 244 distinct GPs were identified and mapped to CEFR levels, representing substantial but selective coverage of the consolidated A1–B2 descriptor space (602 descriptors in total: 73 at A1, 175 at A2, 201 at B1, and 153 at B2). In other words, the corpus realizes just over two-fifths of the available EGP inventory. Table 2 summarizes the distribution of these GPs by CEFR level and educational stage, with percentages indicating the proportion of descriptors covered within each band.

Table 2

Distribution of GPs and EGP Descriptor Coverage by CEFR Level and Educational Stage

CEFR Level	Primary School (CEFR A1–A2 Stage)	Junior High (CEFR A2–B1 Stage)	Senior High (CEFR B1–B2 Stage)
A1 (Beginner)	71 (97.3%)	0	0
A2 (High-Beginner)	35 (20.0%)	74 (42.3%)	0
B1 (Intermediate)	0	58 (28.9%)	89 (44.3%)
B2 (High-Intermediate)	0	0	83 (54.2%)

Notes. Counts reflect distinct GPs explicitly targeted as grammar at each stage, de-duplicated within stage but not across stages. The same GP appearing in multiple volumes at the same stage is counted once for that stage; if it is later re-taught as new grammar content at a higher stage, it is counted again at that higher stage. Percentages in parentheses indicate coverage relative to the full EGP descriptor inventory at each CEFR level (A1: 73 descriptors, A2: 175, B1: 201, B2: 153).

The multi-rater annotation procedure yielded high reliability. As detailed in Reliability and Validity Measures, inter-rater agreement on

CEFR level assignment, GP identification, and recycling/complexity-gain coding fell in the substantial to almost perfect range based on blind, independent coding and was further corroborated by an external expert audit of a stratified sample of entries. These indices, together with documented consensus procedures, indicate that the stage- and series-level contrasts reported in later sections are unlikely to be artifacts of coding noise.

Regarding corpus composition by stage, primary textbooks collectively contributed 106 distinct GPs (predominantly A1–A2 structures), junior-high textbooks 132 GPs spanning A2–B1, and senior-high textbooks 172 GPs including all B2 items in the corpus. Because GPs are de-duplicated within but not across stages, these stage totals exceed the overall set of 244 distinct GPs: they capture the cumulative expansion of the grammar syllabus over time rather than a simple partition of the inventory. Differences in coverage and sequencing across stages and series are examined in the subsequent analyses.

CEFR Stage Alignment of Grammar Distribution

Building on Table 2, which summarizes CEFR distributions across stages, Table 3 further breaks down GPs by stage and textbook series. At this stage level, each GP is counted once in any stage where it is explicitly targeted as grammar, even if it appears in several volumes at that stage; if the same GP is later re-taught as new content at a higher stage, it is counted again at that higher stage. For instance, if *be going to* is presented as a grammar target in multiple primary volumes, it contributes a single A2 count in the primary column of Table 2, and if it is subsequently reintroduced as new grammar content in junior high, it is counted once more in the junior-high column. Within this framework, the overall trajectory of grammar input rises with schooling, though overlaps remain.

Table 3

Number of GPs Covered, by Stage and Textbook Series

Stage	Series (Publisher Code)	No. of Grammar Points Identified *n* (%)
Primary (Grades 3–6)	P-KH	98 (92.5%)
	P-HL	96 (90.6%)
	P-HE	79 (74.5%)
Junior High (Grades 7–9)	J-KH	128 (97.0%)
	J-NY	109 (82.6%)
	J-HL	115 (87.1%)
Senior High (Grades 10–12)	S-SM	162 (94.2%)
	S-LT	158 (91.9%)
	S-HL	151 (87.8%)

Notes. Counts reflect distinct GPs per series across all volumes, de-duplicated within series (within-series recurrences counted once). Percentages indicate coverage relative to the stage-level GP inventory (Primary: 106 distinct GPs; Junior High: 132; Senior High: 172). Totals are not mutually exclusive across series and therefore do not sum to the corpus-wide unique total (244). Values index coverage presence rather than instructional depth; see Cross-Version Differences in Grammar Sequencing for qualitative contrasts. All series are the latest post-2019 curriculum editions.

At the primary level, coverage is overwhelmingly concentrated in the lower band. Seventy-one A1 items cover foundational structures such as the simple present and past, personal and demonstrative pronouns, number marking, high-frequency prepositions, imperatives, and *can* for ability. These 71 items represent almost all of the 73 consolidated A1 descriptors (about 97%) in the EGP inventory. A further 35 A2 items, appearing mainly in upper primary grades, include present continuous, *be going to*, and limited gerund/infinitive uses in fixed expressions, corresponding to about 20% of the 175 A2 descriptors. Taken together, primary textbooks instantiate 106 of the 248 possible A1–A2 units (approximately 43%), indicating a fairly comprehensive treatment of basic forms but a selective introduction of higher-beginner structures.

In the middle years (junior high), the scope broadens to 74 A2 items and 58 B1 items. These figures correspond to approximately

42% of the A2 descriptor set and 29% of the 201 B1 descriptors. Early units reinforce A2 forms where prior coverage was uneven (for example, comparatives and superlatives, past continuous, and zero conditionals). Later units shift toward B1-level constructions, introducing defining relative clauses, first conditionals, simple passives, modals for obligation and advice (*should, must*), and basic reported speech. At this stage, A1 content rarely appears as genuinely new material, functioning primarily as review, whereas B2-level grammar is not yet systematically foregrounded. Overall, junior-high textbooks cover 132 of the 376 possible A2–B1 descriptors (about 35%), signaling a transition from high-beginner to intermediate grammar within a still selective subset of the available space.

By the final stage (senior high), the focus moves predominantly to B1–B2 structures. Eighty-nine B1 items and eighty-three B2 items are explicitly taught, covering roughly 44% of the B1 descriptors and 54% of the B2 descriptors in the consolidated EGP inventory. B1-level treatments deepen earlier topics such as noun clauses, non-defining relative clauses, modals expressing certainty and speculation, and more complex uses of the perfect aspect. Selected B2 forms enter at this stage, including second and third conditionals, inversion after negative adverbials, cleft sentences, and more sophisticated passives (for example, passive infinitives and gerunds). Because the present analysis is restricted to A1–B2, potential C1/C2-oriented language is not considered. Across the B1–B2 space as a whole, senior-high textbooks account for 172 of 354 possible descriptors (approximately 49%), suggesting that advanced coverage is substantial but still partial.

To examine whether the distribution of CEFR levels differs systematically across stages, a chi-square test of independence was applied to the four-level (A1–B2) by three-stage table derived from Table 2. No cells were too sparse for the test; all expected frequencies were at least 5, so the standard chi-square approximation was appropriate. The association between stage and level was significant, $\chi^2(6, N = 410) = 461.51$, $p < .001$, with Cramér's $V = .75$, indicating a strong stage-related shift from lower (A1–A2) to higher (B1–B2) levels. As a robustness check, levels were also collapsed into a lower band (A1–A2) versus a higher band (B1–B2). The resulting 2 × 3 table likewise showed a strong association, $\chi^2(2, N = 410) = 277.98$, p

$< .001$, Cramér's $V = .82$, confirming that the concentration of A1–A2 items in primary and of B1–B2 items in senior high is not an artifact of sparse cells or a particular banding scheme.

In summary, the stage-wise distributions align with the intended progression: primary textbooks concentrate on A1 forms with limited A2 extensions, junior high consolidates A2 and begins to populate the B1 space, and senior high completes much of the B1 band while selectively covering B2 descriptors. At the same time, the coverage ratios (approximately 43% of A1–A2, 35% of A2–B1, and 49% of B1–B2) indicate that, within each target band, only a subset of the EGP inventory is realized as explicit textbook grammar content. Subsequent sections examine how this partial coverage interacts with cross-series variation, recycling patterns, and the timing of early or isolated introductions.

Cross-Version Differences in Grammar Sequencing

Stage-wise coverage patterns were further unpacked by comparing the three series within each educational level (Table 3). Although all publishers work within the same national curriculum guidelines, the series differ measurably in both the number and the selection of GPs, with implications for how consistently learners encounter the consolidated A1–B2 grammar space.

At the primary level, Series P-KH and P-HL are comparable in scope, each introducing just under 100 distinct GPs across eight volumes (98 and 96, respectively), whereas P-HE introduces 79. Given that the primary corpus as a whole contains 106 distinct GPs, this implies that P-KH and P-HL each cover approximately 91–92% of the available primary-level GP inventory, while P-HE covers around three quarters (about 75%). P-HE omits several foundational A1 structures that appear in P-KH and P-HL (e.g., existential *there is*/*there are*; *have got* for possession). In contrast, P-KH extends further into A2 by the end of the stage, introducing *be going to* and frequency adverbs, whereas P-HL is similar in overall quantity but slightly less expansive at A2. Many of these higher-level items in P-KH are introduced late in the stage and receive limited practice, resulting in brief exposure. Overall, P-HE implements a more

conservative sequence that remains largely within A1 and defers a substantial subset of A2 topics to later schooling. The net effect is that primary learners in different series can differ by roughly 15–20 GPs in cumulative exposure, equivalent to about 18% of the stage-level GP inventory, even before entering junior high.

At the junior high level, divergence in density persists. Series J-KH contains 128 GPs, J-NY 109, and J-HL 115, yielding a spread of 19 items, equivalent to approximately 14% of the 132-GP junior-high inventory. In proportional terms, J-KH covers nearly all of the junior-high GPs identified in the corpus (about 97%), J-HL covers around 87%, and J-NY about 83%. J-KH's larger inventory incorporates several B1-oriented structures that are absent from J-NY and J-HL (e.g., stricter obligation with *have to*; degree adverbs such as *absolutely* or *fairly*), although some of these additions receive minimal recycling. J-HL, while slightly leaner overall, devotes more space to systematic reinforcement of core A2 content, particularly through revisiting zero conditionals across multiple grades. This pattern is consistent with a less-is-more approach. J-NY falls between these two profiles in both overall load and the balance between consolidation and extension. Thus, exposure to intermediate-level forms at the same nominal stage of schooling varies systematically with series choice, and the size of the spread is large enough to affect the breadth of learners' cumulative experience with A2–B1 grammar.

In senior high school, all three series surpass 150 GPs in total but diverge most clearly in the scope and treatment of B2-level grammar. S-SM offers the broadest advanced inventory (162 GPs in total), S-LT slightly fewer (158), and S-HL the fewest (151). Relative to the 172 distinct GPs observed at this stage, these totals correspond to coverage of roughly 94%, 92%, and 88% of the senior-high inventory respectively. All three series cover the core B1 topics, including defining versus non-defining relative clauses, major modal distinctions, and passives across tenses. However, they differ in how frequently and in what contexts these items are recycled. More importantly for advanced coverage, S-SM explicitly teaches a wider range of B2 forms with dedicated practice (e.g., third conditionals, inversion after negative adverbials such as *Seldom have I seen …*). S-HL omits these constructions altogether, and S-LT includes them only

marginally (for example, in reading sidebars) without explicit GP status or systematic practice. Consequently, even though all senior-high series appear dense in absolute GP count, learners' opportunities to engage with genuinely B2-level constructions vary considerably by series.

Because descriptive differences among series were largest for advanced grammar, formal cross-series comparisons focused on B2 coverage in senior high. To complement these descriptive contrasts, a simple one-way ANOVA was conducted on descriptor-level coverage of B2 items in senior-high textbooks, treating each B2-mapped GP as a binary outcome (present vs. absent) for each series. The series effect did not reach conventional significance, $F(2, 237) = 1.34$, $p = .26$, $\eta^2 = .01$, but the pattern was consistent with the descriptive results in that the most expansive series covered a somewhat larger subset of B2 descriptors than the leanest series, reinforcing the practical concern that opportunities to engage with advanced grammar are unevenly distributed across senior-high materials.

Taken together, the cross-version discrepancies in Table 3 are quantitatively substantial and qualitatively consequential. Within a given stage, the difference between the densest and leanest series ranges from about 6% to almost 18% of the stage-level GP inventory, and advanced B1–B2 constructions are unevenly introduced and practiced across senior high materials. As a result, students completing the same stage, or even the full K-12 pathway, may emerge having worked through markedly different grammar syllabi. Some will have been explicitly taught a broad range of higher-level constructions, whereas others may only encounter them incidentally in texts or on high-stakes examinations. These patterns underscore the need for clearer, CEFR-referenced guidance on grammar sequencing to minimize unwarranted cross-series variation in learners' cumulative exposure.

Recycling and Vertical Progression Patterns

To gauge cumulative learning, the analysis examined how often GPs were recycled across stages and whether such recycling involved increased complexity. Two dimensions were distinguished: (1) items

exclusive to a single stage (never revisited as target grammar later) and (2) items repeated as explicit target grammar at a subsequent stage. Among the latter, complexity gain was operationalized as any later occurrence that introduced additional formal complexity, extended functions, or more demanding discourse contexts, rather than simply restating the same use. Table 4 summarizes these recurrence types and their frequencies.

Table 4

Recurrence of GPs Across Stages

Recurrence Pattern	Number of Items
Exclusive to one stage (not recycled)	43
Repeated across multiple stages (total)	42
– with increased complexity upon repetition	9
– without added complexity (redundant)	33

Vertical progression can only be meaningfully assessed for items that could, in principle, reappear at a later stage, so the analysis focused on GPs first introduced before senior high. Of the 244 GPs identified in the corpus, 85 met this criterion: 43 items that appeared exclusively within a single stage and 42 items that were explicitly retaught as target grammar in at least one subsequent stage (Table 4). Within this verticality-eligible subset, exclusive and repeated items were roughly balanced. Viewed against the full set of 244 GPs, each group accounts for about 18% of the inventory, indicating that many items that might support cumulative development are either introduced once or retaught only within the same stage.

A closer look at the repeated items shows that genuine vertical progression is relatively rare. Only 9 of the 42 repeated GPs (≈21%) showed clear complexity gain in later presentations, and such vertically aligned cases constitute under 4% of all GPs (9/244 ≈ 3.7%). These instances typically involved an initial, simpler realization followed by a later-stage treatment that added structural or functional demands. For example, comparative adjectives were first introduced in junior high with simple phrasal patterns (e.g., *X is taller than Y*) and subsequently revisited in senior high with clausal comparison or

ellipsis (e.g., *X is taller than Y is*). By contrast, 33 of the 42 repeated items (≈79% of the repeated set; ≈13.5% of the full inventory) were horizontally redundant: they were retaught at a later stage with essentially the same formal scope and usage conditions as at first introduction. In several cases, repetition also reflected inconsistent sequencing across publishers rather than intentionally designed spiraling; for example, *be going to* (A2) is taught in primary Grade 5 in one series but not until junior high in another, so the pooled corpus shows cross-stage recurrence without added complexity.

Stage-related shifts in textbook focus were examined using per-GP counts. For each of the 73 A1-level GPs, the number of textbooks at each stage in which it was assigned explicit GP status was counted. A one-way ANOVA on these counts indicated substantial stage-based differences, $F(2, 70) = 12.47$, $p < .01$, partial $\eta^2 = .26$, with post-hoc tests showing that A1 items were targeted in significantly more primary textbooks than in either junior- or senior-high textbooks (both $p < .01$). A parallel ANOVA on per-GP counts for the 83 B2-level GPs yielded a complementary pattern at the upper end of the scale, $F(2, 80) = 10.29$, $p < .01$, partial $\eta^2 = .20$: senior-high textbooks targeted B2-level GPs in significantly more volumes than junior-high textbooks (which contained virtually none; $p < .01$). Stage-level shifts in CEFR profiles were further examined using a 4 (A1–B2) × 3 (primary, junior, senior) contingency table ($N = 244$). Because some expected cell counts were below 5, Fisher's exact test and log-linear models were used instead of a standard chi-square; both approaches indicated a statistically significant, non-uniform distribution of grammar levels across stages (Fisher's exact $p < .001$; Cramér's $V = .30$), consistent with the descriptor-level pattern in CEFR Stage Alignment of Grammar Distribution: lower-level items cluster in earlier stages and higher-level items in later stages, with an over-representation of A2-level items in primary and relatively limited B2 coverage even by senior high.

To examine recycling patterns more directly, a logistic GLMM was fitted with cross-stage recycling (0 = not recycled; 1 = recycled) as the dependent variable, CEFR level and series as fixed effects, and random intercepts for stage (primary, junior high, senior high). The model revealed a strong main effect of CEFR level on the likelihood

that a GP would be recycled (Wald $\chi^2(3) = 27.46$, $p < .001$). Intermediate-level items (A2–B1) were substantially more likely to be retaught in a later stage than either very basic (A1) or advanced (B2) items. B1-level GPs showed the highest recycling rate, with estimated odds of cross-stage repetition almost an order of magnitude higher than for A1 items (odds ratio = 9.03, $p < .001$), while A2-level GPs also had higher odds of recycling than A1 (odds ratio = 4.12, $p < .01$). By contrast, A1-level GPs, which represent foundational structures concentrated in primary school, had a very low probability of being retaught as new grammar in later stages (baseline recycling probability below 10%). B2-level GPs were least likely to be recycled, as they appeared almost exclusively in senior-high textbooks with no later stage available for further recycling These results show that cross-stage recycling is most common for intermediate grammar, especially B1, whereas beginner and advanced bands are largely confined to the initial and final stages of schooling, respectively.

The same GLMM also identified significant differences across textbook series in recycling patterns (Wald $\chi^2(2) = 6.32$, $p = .042$). Within each stage, the most expansive series, defined as the one with the largest number of distinct GPs (Series A), exhibited the highest proportion of recycled items. The more conservative or minimalist series (Series B) showed the lowest proportion, with the remaining series (Series C) falling between these extremes. Post-hoc comparisons indicated that the odds of a GP being recycled were approximately twice as high in Series A as in Series B ($p < .05$), while differences involving Series C were less consistent once CEFR level was controlled. This pattern aligns with the descriptive findings reported in Cross-Version Differences in Grammar Sequencing: series that introduce a broader range of GPs earlier create more overlap of content across stages and hence more opportunities for cross-stage repetition, whereas minimalist series tend to adopt a delay-and-defer approach, covering many GPs only once at a later grade.

Overall, the mixed-effects results converge with the ANOVA and contingency analyses reported above. The low recycling odds for A1 items and near-zero odds for B2 items mirror the stage-wise concentration of A1 content in primary and B2 content in senior high, while the elevated recycling probabilities for B1 (and, to a lesser

extent, A2) GPs match the descriptive pattern that intermediate grammar spans junior and senior high. Taken together, these findings suggest that although the overall distribution of GPs shifts upward in CEFR level across stages, the mechanisms that support cumulative growth are uneven. Within the subset of items most relevant to verticality, many appear only once in the K-12 sequence or are retaught without added complexity, and only a small minority display clear complexity gain. A more strongly spiral curriculum would entail more systematic, stage-sensitive reintroduction of forms with expanded functions, denser discourse contexts, or more integrative tasks, rather than predominantly horizontal recycling of earlier treatments.

Early and Isolated Grammar Introductions

The analysis also identified sequencing misalignments of two main types: early introductions (structures appearing at a lower stage than pedagogically expected) and isolated introductions (advanced structures appearing late without preparatory groundwork or follow-up). Such cases signal points of strain where learners may struggle if content is introduced prematurely, or remain underprepared if difficult items arrive without sufficient buildup.

Premature introductions in primary: Although primary textbooks overwhelmingly target A1-level GPs, several A2-level structures appeared earlier than developmental sequences in SLA research would typically predict. Three A2-level items were coded as clearly premature: the present perfect for experience, embedded *wh*-questions (e.g., *I know what you did.*), and *be going to* (future). These structures surfaced in upper primary materials, with the present perfect and embedded questions introduced by the end of Grade 6 in two series (P-KH, P-HL), and *be going to* taught as early as Grade 5 in P-KH. Taken together, these three items represent approximately 9% of the A2 descriptors covered in primary (3 out of 35) and about 1% of all GPs in the corpus (3 out of 244). All three appeared in single units with little systematic recycling in junior high, implying that exposure was brief and unsupported by a sustained instructional trajectory.

Isolated advanced items in senior high: At the other end of the

sequence, several B2-level items appeared in ways that were judged isolated, in the sense that they lacked scaffolding from related, simpler structures. Two B2 constructions were coded in this way. First, third conditionals were introduced in Grade 11 in Series S-SM, presupposing secure mastery of second conditionals (B1) and comfort with past-modal verb forms; yet Series S-HL did not explicitly cover second conditionals at all, and treatment in S-LT was limited. Second, inversion after negative adverbials (e.g., *Hardly had I arrived …*) appeared in the final year of S-SM but was absent from S-LT and S-HL. For learners using S-SM, these topics amounted to one-off lessons on specialized B2 forms; for others, they were not explicitly taught. Numerically, these two structures account for about 2% of the B2 descriptors covered in senior high (2 out of 83) and less than 1% of all GPs in the corpus (2 out of 244). When the advanced items coded as premature and isolated are pooled, they constitute around 4% of the A2/B2 descriptors actually targeted at these stages (5 out of 35 A2-in-primary plus 83 B2-in-senior).

Omitted advanced descriptors: Beyond these early and isolated cases, several grammar features within the nominally appropriate stage bands did not appear at all in the textbooks. Of the 201 B1 descriptors in the EGP, only 89 (44%) were represented in any senior-high textbook, and 58 (29%) had already been introduced by the end of junior high; thus around 56% of B1 descriptors never appeared in the K-12 sequence. For B2, 83 of 153 descriptors (54%) were covered in senior high, leaving about 46% of B2 descriptors unaddressed. Pooling B1 and B2 together, this corresponds to just over half of the advanced descriptor inventory (≈ 51%) never being explicitly targeted in any textbook. While covering every descriptor is neither necessary nor practical, the scale of these omissions indicates that many potentially important structures remain outside the taught syllabus despite their relevance to academic reading and formal writing at B2 and their likely presence in high-stakes exams. Examples include ellipsis in comparative clauses (e.g., *Mary is taller than John (is).*) and past modal constructions for speculation (e.g., *He might have forgotten.*).

Taken together, these quantitative and qualitative patterns indicate a selective approach to advanced grammar. On the one hand, only a

small fraction of the advanced items actually taught at primary and senior levels were coded as clearly mis-sequenced (approximately 9% of A2 items in primary and 2% of B2 items in senior high, or about 4% of these advanced targets overall). On the other hand, these misalignments occur against a backdrop where roughly half of the B1–B2 descriptor pool is never introduced at all. In other words, problems of premature or isolated introduction coexist with a much larger issue of systematic omission. The combined effect is a progression in which some learners encounter demanding forms too early or without scaffolding, while a substantial proportion of potential advanced grammar remains outside the core school curriculum. These mismatches between an ideal, scaffolded progression and the actual sequencing are further considered in the Discussion in relation to developmental appropriateness and curriculum planning.

DISCUSSION

Stage-aligned Progression with Uneven Completeness

At a broad level, the corpus exhibits a stage-sensitive progression: primary materials focus on A1 GPs, junior high consolidates A2 and begins B1, and senior high emphasizes B1 with selective B2 coverage. This contour aligns with CEFR-informed curriculum goals, with CEFR and EGP levels treated as descriptive reference points rather than prescriptive standards, and is consistent with SLA accounts that recommend postponing more complex clause structure and modal–aspectual systems to later stages.

This progression is, however, unevenly complete. Some A2-profiled features (e.g. *be going to* for future intention, the present perfect for experience) appear in primary textbooks but receive little systematic, spaced recycling, limiting opportunities for younger learners to consolidate them. At the other end, B2-related constructions, while more numerous in senior-high books than at earlier stages, are still relatively limited in scope and are often confined to isolated lessons, making repeated encounters with upper-intermediate patterns in varied contexts unlikely. The overall direction

of progression is therefore developmentally plausible, but the intensity and consistency of exposure to particular GPs are insufficient to guarantee sustained grammatical development.

Cross-series Variability and Equity of Grammatical Opportunity

Even under a single national curriculum, the three textbook series diverge in both the selection and density of targeted GPs. At the primary level, one series omits foundational structures (e.g., basic existential constructions), whereas another reaches ahead into A2-type future expressions. At junior high, one series broadens coverage to a range of B1-related patterns while another focuses more narrowly on consolidating A2. By senior high, advanced items such as third conditionals and inversion after negative adverbials are systematically taught in one series but are absent or only glossed in others. Learners thus follow distinct grammatical input trajectories, with some cohorts receiving earlier or richer exposure to key constructions than others.

Some variability is to be expected and may reflect different pedagogical priorities, such as broader exposure versus deeper mastery. The divergence observed here, however, raises equity concerns. When high-value GPs that enable important communicative functions or are associated with higher proficiency bands in CEFR-informed curricula are omitted, marginalized, or only superficially treated, learners are offered narrower developmental opportunities, especially in systems where high-stakes assessments presuppose familiarity with such forms. Experience from other CEFR-referenced reforms suggests that broad policy signals without an explicit grammar scope-and-sequence invite heterogeneous interpretations by textbook authors, resulting in uneven coverage of core grammatical competencies. A published core inventory of GPs, specifying which structures should be introduced and revisited at each stage, could help stabilize opportunity across series while still allowing diversity in themes, texts, and task design.

Vertical Coherence: Recycling, Complexity Gain, and Recurrent Failure Modes

A spiral curriculum assumes that when grammatical forms

reappear at later stages, they do so with increased functional or structural complexity. Examples include the progression from basic comparatives to clausal comparison and the extension of the present perfect from simple experiential uses to discourse-organizing functions. In the present corpus, cross-stage recurrence of GPs is common, but meaningful complexity gain is rare. Many later appearances closely replicate the initial presentation, reteaching the same rules with similar examples and exercise formats rather than extending meanings or uses. Of the 42 GPs that were explicitly retaught as target grammar at a subsequent educational stage, only 9 (about 21%) showed a clear increase in structural or functional complexity, so genuine spiraling up is the exception rather than the norm. From a skill-acquisition perspective, repeated encounters that do not increase cognitive or communicative demands are likely to stabilize primarily declarative knowledge, offering only partial support for proceduralization and increasing the risk of developmental plateaus in learners' use of those forms. The GLMM results reported in Recycling and Vertical Progression Patterns corroborate this pattern, showing that cross-stage recycling is statistically concentrated in intermediate CEFR bands (especially B1) and occurs more frequently in expansive series than in minimalist ones, while beginner-level (A1) and advanced-level (B2) items are seldom retaught across stages.

The analysis points to three recurrent failure modes. First, mis-timed introduction arises when A2-related GPs are introduced in early primary without sustained, age-appropriate recycling, or when B2-related patterns appear late and in isolation, leaving learners either under-prepared or with too little time to practice them across contexts. Second, redundant reteaching occurs when higher-level units revisit a GP with almost identical explanations and task types to those used at first introduction; such repetition may be intended as review or as insurance against uneven prior coverage, but without added functions or processing demands it contributes little new learning. Third, coverage gaps occur where GPs indexed at B1–B2 in CEFR-related descriptor systems (e.g., modals for past speculation, more complex comparative structures, cleft sentences) are largely absent from a series; when such features are also referenced in curriculum documents or implicit in examination demands, uneven coverage

across series translates into unequal access to the grammatical resources that matter for both classroom communication and high-stakes assessment. Taken together with the mixed-effects findings, these patterns indicate that vertical coherence is weakest precisely where sustained, level-sensitive recycling would be most pedagogically valuable.

Standards Architecture: Operationalizing CEFR-T with an EGP-Linked Checklist

The emerging CEFR-T framework in Taiwan can support more coherent grammar progression only if it specifies not just broad outcome bands (e.g., "A2 by the end of primary") but also concrete stage-specific entry points and growth targets for key GPs. Within such an architecture, CEFR-T would serve as a national point of reference where global CEFR descriptors are interpreted for the local context, translating abstract proficiency bands into decisions about when to first introduce and how to later extend particular grammatical resources. A coherent plan would (a) assign each priority GP a stage-appropriate first introduction, (b) mandate at least one planned revisit where that GP is taken up with added functional or structural complexity, and (c) publish a core inventory of essential GPs per stage, with options for enrichment beyond the core. Such a shared inventory would shift textbook review from ad hoc repetition to intentional spiraling and would help limit unnecessary divergence across series while preserving room for thematic and methodological variation.

Building on the evaluative dimensions summarized in Table 1, the EGP offers a corpus-based foundation for operationalizing this architecture through a scope–level–recycling (SLR) checklist. In an EGP-linked SLR grid, each targeted GP can be documented with (a) the entry-level EGP descriptor, (b) the descriptor that defines the intended growth at recycle, (c) the unit and page where these uses are modeled, and (d) an example activity or assessment prompt that elicits the more advanced use. Developers can employ such a checklist to audit coverage and planned complexity gain during writing; reviewers can use it to evaluate draft materials against CEFR-T-related guidance. By making omissions and missed opportunities for development

visible at the design stage, an SLR checklist provides a practical bridge between descriptive benchmarks (CEFR/EGP) and the concrete task of constructing a coherent grammar strand, without turning reference levels into rigid prescriptions.

The checklist proposed here is, however, a conceptual tool rather than a validated solution. Its feasibility and impact still need to be examined through pilot use in publisher workflows or school-level materials review. Such evaluation would help determine whether the checklist supports more consistent grammar grading and identify any additional workload or training needs.

Implications, Limitations, and Future Directions

Pedagogically, the findings suggest that early appearances of more advanced forms need not be viewed as errors in themselves. When such items are revisited in planned, spaced ways and embedded in meaning-rich tasks, initial exposure can serve as a productive preview of language slightly above the prevailing level. Likewise, repeated treatment of the same GP across grades can support development if later encounters expand functional range or discourse integration. For example, a review unit on the past continuous could move beyond rule rehearsal to narrative construction, stance-taking, or exploration of hypothetical scenarios, thereby increasing the communicative load of a familiar form. In practice, however, many recurrences are horizontally repetitive and substantial portions of the advanced descriptor space are not addressed at all. Expecting individual teachers to systematically compensate for these structural gaps by redesigning tasks is unrealistic, especially in large, exam-oriented classes. System-level improvements in materials design, CEFR-T guidance, and professional learning are therefore essential if grammar pathways are to offer coherent, cumulative opportunities for growth across the K-12 span.

These implications are qualified by several limitations. First, the analysis was restricted to student textbooks. Workbooks, teachers' guides, digital supplements, classroom interaction, and learner performance were not systematically examined. The coverage figures and gaps reported here should thus be interpreted as upper-bound

estimates of the grammar input encoded in core student materials. Some GPs absent from the student books may be addressed in ancillary resources, partially offsetting the omissions identified, whereas not all textbook content is necessarily taught, practiced, or retained. The available data do not permit estimating how often supplementary materials systematically offset the documented omissions; even substantial supplementary use would only partly mitigate the absence of advanced GPs from the core student texts. Any comments about possible consequences for learning, including the risk of developmental plateaus or uneven preparation for upper-intermediate use, are therefore theory-informed interpretations of input patterns rather than empirical claims about attainment. These interpretations require confirmation against classroom and learner data. Second, the study focuses on a single CEFR-referenced curriculum and textbook system. The patterns observed reflect Taiwan's specific configuration of CEFR-T-related policy, textbook approval procedures, and high-stakes examinations, and cannot be assumed to hold in other contexts where curricular aims, instructional traditions, and assessment regimes differ. A further limitation is that the analyses differentiated between coverage gaps and redundant reteaching but did not statistically model their association across series (e.g., whether series with more gaps also display more redundancy), so any correlations among these failure modes remain to be investigated in future work.

Against this backdrop, several lines of further work are warranted. At the classroom level, textbook-based grammar maps should be triangulated with lesson observations, teacher interviews, assessment artifacts, and learner corpora to establish whether the misalignments and omissions documented here correspond to persistent learner difficulties, fossilized errors, or compensatory practices. Such research would also illuminate how teachers interpret and adapt textbook sequencing in real instructional settings and how much scope they perceive for creating their own form–meaning trajectories. At the system level, periodic audits of new textbook editions could monitor whether more intentional spiraling and clearer growth targets for key GPs gradually replace the current mix of selective coverage and ad hoc review, and whether advanced grammar coverage becomes

more even across series and stages. Such audits would be particularly relevant as CEFR-T specifications become more fine-grained. A further priority is to compare textbook-based grammar input with the demands of high-stakes assessments such as Taiwan's General Scholastic Ability Test (GSAT), in order to check that students using different series enjoy broadly comparable opportunities to prepare for exam-relevant structures. Finally, the analytic framework developed here can be applied to other CEFR-informed systems to support cross-context comparisons of graded grammar input. Such comparative work would help distinguish issues that are widespread in CEFR-referenced textbook regimes from those that are specific to the Taiwanese implementation and would inform ongoing efforts to use CEFR as a flexible point of reference for supporting learning and learner development rather than as a rigid script for instruction.

CONCLUSION

This study has documented how grammar is staged and recycled across Taiwan's K-12 EFL student textbooks in a CEFR-informed policy environment. Using the EGP as a corpus-based reference and drawing on SLA perspectives, it mapped 244 textbook-intended GPs across three schooling stages and three approved series, treating CEFR-related levels as descriptive indicators of input bands and developmental potential rather than prescriptive targets. The analysis shows that the materials broadly move from A1-oriented content in primary school to A2–B1 in junior high and B1–B2 in senior high, consistent with national curriculum aims.

Viewed through a developmental lens, however, this progression is selective and uneven. Many core forms are well covered, yet a sizable proportion of potentially important B1–B2 descriptors never enters the taught syllabus, cross-series differences in scope and timing are substantial, and most cross-stage recurrences involve simple repetition rather than clear expansion in function or structural complexity. Advanced items are sometimes introduced briefly or late, while some demanding forms appear earlier than is supported by sustained practice. Together, these patterns suggest that the current

textbook system offers uneven opportunities for learners to extend their grammatical resources over time, even though the general direction of level progression is upward.

Conceptually and methodologically, the study illustrates how CEFR-related resources and learner-corpus-based descriptors can be used to audit grammar strands in school materials. By linking level information to processability, practice conditions, and opportunities for form–meaning use, it offers a stage-sensitive template for examining how policy documents, textbooks, and developmental principles interact in a CEFR-informed environment. The proposed scope–level–recycling checklist points to one practical way in which CEFR-T and the EGP could be integrated into curriculum and materials work to support more transparent decisions about coverage, sequencing, and vertical progression.

For Taiwan and comparable EFL systems, these findings underscore the value of sustained, data-informed collaboration among policymakers, textbook writers, teacher educators, and researchers. Strengthening the grammar strand within an action-oriented, learner-focused curriculum will require clearer stage-level guidance, more consistent planning of complexity gains across grades, and complementary evidence from classrooms and learner performance. The analytic framework developed here provides a baseline against which future textbook revisions, CEFR-T implementation efforts, and cross-national comparisons can be evaluated, with the broader goal of supporting equitable, long-term growth in learners' grammatical and communicative repertoires.

REFERENCES

Ahmad Afip, L., Hamid, M. O., & Renshaw, P. (2019). Common European framework of reference for languages (CEFR): Insights into global policy borrowing in Malaysian higher education. *Globalisation, Societies and Education*, *17*(3), 378-393. https://doi.org/10.1080/14767724.2019.1578195

Albahouth, K. (2024). *The impact of a pedagogic intervention on learner engagement in Saudi EFL grammar classes: A focus on flipped and interactive pedagogy* (Doctoral dissertation). University of Southampton.

Alsalem, M. S. (2020). *Language teacher cognition on English grammar assessment: Investigating EFL teachers' beliefs and practices regarding classroom-based assessment in public higher-educational institutions in Saudi Arabia* (Doctoral dissertation). University of Essex.

Aman, Z. B. (2021). *Teachers' perceptions toward the implementation of CEFR-aligned textbook in secondary school* (Unpublished master's project report). Universiti Teknologi Malaysia.

Azaha, N. R., Li, Y., Nimehchisalem, V., & Mukundan, J. (2024). Are we presenting vocabulary items properly? The case of the CEFR-aligned Malaysian Form 4 English Language Textbook. *Open Journal of Modern Linguistics*, *14*(6), 1040-1059. https://doi.org/10.4236/ojml.2024.146055

Baron, P. A., & Papageorgiou, S. (2014). *Mapping the TOEFL® Primary™ Test onto the Common European Framework of Reference*. (Research Memorandum No. RM-14-05). ETS.

Bell, P., & Payant, C. (2024). Comprehensible input in language instruction: Approaches and techniques. In W. Wong & J. Barcroft (Eds.), *The Routledge handbook of second language acquisition and input processing* (pp. 315-328). Routledge. https://doi.org/10.4324/9781003041412-28

Burton, G. F. (2020). *The canon of pedagogical grammar for ELT: A mixed methods study of its evolution, development and comparison with evidence on learner output*. (Unpublished PhD thesis). Mary Immaculate College, University of Limerick.

Chali, G. T. (2020). *An investigation of EFL teachers' beliefs on teaching grammar and reading and their practices in secondary schools in Ethiopia* (Doctoral dissertation). Ghent University, Belgium.

Chang, L. Y. H. (2020). Taiwanese EFL language teachers' beliefs and actual practices related to learner autonomy. *TESL-EJ*, *23*(4), 1-15.

Chen, K. Y. (2018). *University English teachers' ideologies in respect of English language policy in Taiwanese higher education* (Doctoral dissertation). University of Southampton, England.

Cheng, Y. S., & Chang, S. C. (2022). Do senior high school English textbooks sufficiently prepare students for the high-stakes college entrance examinations? A corpus-based analysis of text difficulty. *Journal of Textbook Research*, *15*(1), 43-80. https://doi.org/10.6481/JTR.202204_15(1).02

Chiang, H. L. L. (2003). *EFL teachers' beliefs and practices at an exemplary*

Taiwanese elementary school (Doctoral dissertation). The University of Texas at Austin.

Chien, C. W. (2023). Elementary school English teachers' beliefs and classroom practice regarding alternative curricula. *Journal of Language, Identity & Education*, *22*(2), 153-166. https://doi.org/10.1080/15348458.2020.1845701

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment—Companion volume*.

Criado, R. (2016). Insights from Skill Acquisition Theory for grammar activity sequencing and design in foreign language teaching. *Innovation in Language Learning and Teaching*, *10*(2), 121-132. https://doi.org/10.1080/17501229.2015.1090996

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15-32). Routledge. https://doi.org/10.4324/9781315676968-2

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(1), 83-107. https://doi.org/10.2307/40264512

Fives, H., & Buehl, M. M. (2016). Teachers' beliefs, in the context of policy reform. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 114-121. https://doi.org/10.1177/2372732215623554

Freedman, A. (1993). Show and tell? The role of explicit teaching in the learning of new genres. *Research in the Teaching of English*, *27*(3), 222-251. https://doi.org/10.58680/rte199315402

Gass, S. M. (2015). Comprehensible input and output in classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 182-197). Wiley-Blackwell. https://doi.org/10.1002/9781118531242.ch11

Grahn Sherin, J. (2023). *Reviewing the effects of explicit and implicit grammar instruction in second language acquisition: A literature review* [Student thesis, Linköping University].

Hamid, H. A., Fakhruddin, W. F. W. W., Afip, L. A., & Hamid, M. O. (2025). Navigating global standards in local contexts: A systematic review of CEFR-aligned English language curriculum in the ASEAN region. *Journal of Nusantara Studies (JONUS)*, *10*(1), 324-352. https://doi.org/10.24200/jonus.vol10iss1pp324-352

Haney, J. J., Lumpe, A. T., Czerniak, C. M., & Egan, V. (2002). From beliefs to actions: The beliefs and actions of teachers implementing change. *Journal of Science Teacher Education*, *13*(3), 171-187. https://doi.org/10.1023/A:1016565016116

Hong, J. F., Tseng, H. C., Peng, C. Y., & Sung, Y. T. (2020). Linguistic feature analysis of CEFR labeling reliability and validity in language textbooks. *Journal of Technology & Chinese Language Teaching*, *11*(1), 57–83.

Hsieh, I. (2022). *Policy, position, and practice: a linguistic ethnography study of Taiwanese English language teacher identity and agency in multilingual and multimodal times* (Doctoral dissertation). University of Warwick, England.

Im, G. H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: Their limitations and future directions for validation research. *Language Testing in Asia*, *9*(1), 14. https://doi.org/10.1186/s40468-019-0089-4

Jegerski, J. (2021). Krashen and second language processing. *Foreign Language Annals*, *54*(2), 318-323. https://doi.org/10.1111/flan.12557

Jeon, J. H. (2022). A systematic review of CEFR-related research of English education in South Korea. *Journal of Curriculum and Teaching*, *11*(8), 363-375. https://doi.org/10.5430/jct.v11n8p363

Kareema, M. I. F. (2023). *A Rasch approach to validation of CEFR-aligned reading tests for testing undergraduate reading comprehension* (Unpublished doctoral dissertation). International Islamic University Malaysia.

Ke, I. C. (2021). *Globalization and English education in Taiwan: Curriculum, perceptions, and pedagogies*. Routledge.

Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, *19*(2), 142-161. https://doi.org/10.1080/15434303.2021.2023542

Krashen, S. (1992). The input hypothesis: An update. In J. E. Alatis (Ed.), *Linguistics and language pedagogy: The state of the art* (pp. 409-431). Georgetown University Press.

Krashen, S. (1997). The comprehension hypothesis: Recent evidence. *English Teachers Journal (Israel)*, *51*, 17-29.

Krashen, S. (2017). The case for comprehensible input. *Language Magazine*, *7*(1), 1-6.

Kulaklı, E. (2024). *Comparative analysis of the 12th-grade English curriculum and coursebook with the CEFR* (Master's thesis). Bilkent Universitesi, Turkey.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Lichtman, K., & VanPatten, B. (2021). *Was Krashen right? Forty years later*. *Foreign Language Annals*, *54*(2), 283-305. https://doi.org/10.1111/flan.12552

Lin, K. P. (2024). Taiwanese teachers' beliefs toward EFL learner autonomy and their practices in high school. *Journal of Modern Educational Research*, *3*, 10. https://doi.org/10.53964/jmer.2024010

Lin, S. W. (2020). *Analysis of Taiwan's English curriculum and teaching materials from the perspective of English as a lingua franca*. https://doi.org/10.21203/rs.3.rs-42325/v1

Little, D., & Figueras, N. (Eds.). (2022). *Reflecting on the common European framework of reference for languages and its companion volume* (Vol. 104). Multilingual Matters. https://doi.org/10.2307/jj.22679730

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition*, *16*(3), 303-323. https://doi.org/10.1017/S0272263100013103

Lovrić, D. (2024). *Grammar assessment in EFL classrooms* (Doctoral dissertation).

University of Zadar, Croatia.
Lynn, E. M. (2021). *The linguistic nature of ESL/EFL reading textbooks: A corpus analysis* (Doctoral dissertation). Northern Arizona University.
Nagai, N., Birch, G. C., Bower, J. V., & Schmidt, M. G. (2020). *CEFR-informed learning, teaching and assessment*. Springer Singapore.
Negishi, M., Takada, T., & Tono, Y. (2013). A progress report on the development of the CEFR-J. In E. D. Galaczi & C. J. Weir (Eds.), *Exploring language frameworks: Proceedings of the ALTE Kraków Conference, July 2011* (pp. 135-163). Cambridge University Press.
Ng, C. H., Cheung, Y. L., & Zhang, H. (2022). Current trends and future directions in English language assessment practices in the transnational world: A synthesis of recent research. In O. Z. Barnawi, M. S. Alharbi, & A. A. Alzahrani (Eds.), *Transnational English language assessment practices in the age of metrics* (pp. 239-256). Routledge. https://doi.org/10.4324/9781003252382-19
Nguyen, H. T. D. (2024). *From textbook adoption to implementation: Investigating teacher agency amid educational reform in Vietnam* (Doctoral dissertation). University of Illinois at Urbana-Champaign.
North, B., & Piccardo, E. (2023). Aligning language frameworks: An example with the CLB and CEFR. *Language Assessment Quarterly*, *20*(2), 143-165. https://doi.org/10.1080/15434303.2023.2184266
Ohashi, Y., Katagiri, N., & Oshikiri, T. (2023). Vocabulary Analyzer Based on CEFR-J Wordlist for Self-Reflection (VACSR) Version 2. *Asia Pacific Journal of Corpus Research*, *4*(2), 75-87.
O'Keeffe, A. & Mark, G. (2017). The English grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, 22(4), 457-489. https://doi.org/10.1075/ijcl.14086.oke
Pienemann, M. (1998). *Language processing and second language development: Processability theory* (Vol. 15). John Benjamins Publishing. https://doi.org/10.1075/sibil.15
Savski, K. (2020). Local problems and a global solution: Examining the recontextualization of CEFR in Thai and Malaysian language policies. *Language Policy*, *19*(4), 527-547. https://doi.org/10.1007/s10993-019-09539-8
Shao, L. (2019). *Case Studies of English-Medium Instruction in Higher Education: Business Programmes in China, Japan and the Netherlands*. (Unpublished doctoral dissertation). University of Dublin, Trinity College, Ireland. www.tara.tcd.ie/bitstream/handle/2262/88792/Final-E-Thesis%20upload-with-%20transcripts-final%20submission-Lijie-thesis-.pdf
Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263-308. https://doi.org/10.1111/j.1467-9922.2010.00562.x
Sulaiman, F., & Singh, S. K. S. (2021). Grammar profiling: Realignment of CEFR teaching materials and assessments. *International Journal of Research and Innovation in Social Science*, *5*(11), 593-597. https://doi.org/10.47772/

IJRISS.2021.51132

Timperley, H., & Alton-Lee, A. (2008). Reframing teacher professional learning: An alternative policy approach to strengthening valued outcomes for diverse learners. *Review of Research in Education*, *32*(1), 328-369. https://doi.org/10.3102/0091732X07308968

Tono, Y., & Negishi, M. (2020). *Kyozai tesuto sakusei no tame no CEFR-J risosu bukku* [CEFR-J resource book to develop materials and tests]. Taishukan Shoten.

Tseng, C. C. (2019). Senior high school teachers' beliefs about EFL writing instruction. *Taiwan Journal of TESOL*, *16*(1), 1-39. https://doi.org/10.30397/TJTESOL.201904_16(1).0001

Tsui, A. B. M. (2020). Glocalization and grobalization: Critical issues in English language teaching and teacher education in East Asia. In A. B. M. Tsui (Ed.), *English language teaching and teacher education in East Asia: Global challenges and local responses* (pp. 1-36). Cambridge University Press. https://doi.org/10.1017/9781108856218.002

Uri, M. (2021). CEFR in Malaysia: *The appropriacy of English syllabus and assessment, stakeholders' perspectives and challenges.* (Unpublished Ph.D. thesis). Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Malaysia.

Valax, P. (2011). *The Common European Framework of Reference for Languages: A critical analysis of its impact on a sample of teachers and curricula within and beyond Europe* (Doctoral dissertation). University of Waikato, New Zealand.

Vettori, G., Casado Ledesma, L., Tesone, S., & Tarchi, C. (2024). Key language, cognitive and higher-order skills for L2 reading comprehension of expository texts in English as foreign language students: A systematic review. *Reading and Writing, 37*(9), 2481-2519. https://doi.org/10.1007/s11145-023-10479-3

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, X. (2011). *Grammatical development among Chinese L2 learners: From a processability account* (Doctoral dissertation). Newcastle University.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence. *Applied Linguistics*, *8*(2), 95-110. https://doi.org/10.1093/applin/8.2.95

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers of the *Taiwan Journal of TESOL* for their constructive comments, which helped improve earlier versions of this paper.

CORRESPONDENCE

Tzu-Fen Nellie Yeh, Research Center for Translation, Compilation, and Language Education, National Academy for Educational Research, Taipei, Taiwan.
Email address: nellieyeh@gmail.com

PUBLISHING RECORD

Manuscript received: October 14, 2025; Revision received: December 4, 2025; Manuscript accepted: February 20, 2026.