



Lexical Quality Development in Korean EFL Learners' Spelling: A Feature-Based and Distributional Analysis across Grades 4–6

Han Suk Bae (Dong-A University)

Received: 17 April 2026
Revised: 1 May 2026
Accepted: 20 May 2026

Bae, Han Suk. (2026). Lexical quality development in Korean EFL learners' spelling: A feature-based and distributional analysis across grades 4-6. *Modern English Education*, 27, 292-307.

Keywords

Korean EFL student,
English spelling,
lexical quality development,
primary spelling inventory
한국인 영어학습자,
영어 철자,
어휘 표상의 질적 발달,
초등 철자 평가 도구

Han Suk Bae

Assistant Professor
Department of Tourism
Management
Dong-A University
hsbtamu@dau.ac.kr

*This research was supported by
Dong-A University research fund.

Abstract

This study examines spelling development among Korean elementary EFL learners through the Lexical Quality Hypothesis (LQH). A total of 139 students in Grades 4–6 participated in a revised Primary Spelling Inventory. The analyses focused on grade-level differences in spelling accuracy, patterns of feature-level dominance, and the structure of spelling errors, employing type–token ratio and normalized Shannon entropy to measure error diversity and distributional evenness, respectively. The results indicated significant overall improvements in spelling accuracy across grades, although the rate of growth slowed down over time. Feature-level analyses showed that phonological and orthographic features developed concurrently, rather than sequentially, with consonant-based features consistently dominating. In contrast, morpho-orthographic features, particularly inflected endings, exhibited ongoing instability. Error analyses revealed various developmental trajectories, including the expansion and competition of alternative forms, consolidation toward conventional spellings, and instances of reduced variability that suggested premature stabilization of incorrect representations. These findings imply that spelling development is not a straightforward process of error reduction; instead, it is a dynamic process of lexical restructuring characterized by competing representations. The study supports the LQH by illustrating that variability in errors reflects the evolving quality of lexical representations, underscoring the importance of explicit, feature-based spelling instruction in EFL contexts.

INTRODUCTION

There have been changes in the viewpoints regarding children's spelling development. Until the 1960s, spelling skill was considered somewhat dichotomous, classifying children as either good or bad spellers. In this primitive view, children's spelling errors were considered as errors representing the pre-competence stage (Brown, 1990). However, with a growing understanding of children as active and creative language users in the 1970s (Chomsky, 1976; Read, 1973), the view began

to face criticism and became modified with a more elaborate point of view, namely, stage theory (Ehri, 1989; Gentry, 1981; Henderson & Templeton, 1986). After the 1980s, the developmental approach to spelling was prevailing, and opened the way to understand children's spelling errors as the mirror of the student's creative language use. Recent understanding of spelling has changed from the strict viewpoint of the stage model to the strategy model based on in-depth observations of children's invented spellings (Treiman 1993; Treiman & Bourassa, 2000).

Regarding Korean EFL learners, little attention has been given to explore how the learners' spelling skills are developed. Among the dearth, Y. Lee (2007a, 2007b) representing some of the earliest systematic attempts to address this gap, in that the study examined the relationship between English spelling accuracy and spelling sophistication among fourth-, fifth-, and sixth-grade elementary students. The findings showed that spelling accuracy increased with grade level and that the nature of spelling errors shifted from simple phonological substitutions toward the use of more complex orthographic strategies. More recently, S. Lee (2019) examined elementary students' English spelling through an analysis of invented spellings. This study emphasized learners' nonstandard spellings as windows into their underlying hypotheses about English orthography, highlighting how students actively construct and test spelling knowledge.

However, neither line of research fully explicates how distinct components of lexical knowledge—such as consonant representation, vowel pattern selection, digraph processing, and inflectional morphology—are integrated within word-level representations over time. This gap highlights the need for a more fine-grained and theoretically grounded account of spelling development in Korean EFL learners. From the perspective of the Lexical Quality Hypothesis (LQH) (Perfetti, 2007), spelling development can be understood as the gradual integration and stabilization of phonological, orthographic, and morphological information within lexical representations. However, this perspective has rarely been applied to the analysis of Korean EFL learners' spelling development. The present study addresses this need by examining spelling development across grades using both feature-level and distributional analyses.

LITERATURE REVIEW

Spelling Development and the Lexical Quality Hypothesis

Research on children's spelling development has progressively shifted from viewing spelling errors as indicators of deficiency toward understanding them as evidence of developing linguistic knowledge. In the early 1970s, studies by Charles Read and Carol Chomsky paved the way for re-considering children's spelling errors as representations of their creative language use. Their work introduced a notion of "invented spelling", and was named after the finding that children's attempts to spell were processes of inventing their own spelling with systematic logic rather than confusion or misunderstandings (Bear et al., 2020). Charles Read provided examples of consistent logic in children's invented spelling (Read, 1975): vowels are represented with the letter name (e.g., BOT for *boat* and FEL for *feel*) and nasals before consonants are omitted (e.g., WOT for *won't* and BOPY for *bumpy*). Carol Chomsky also provided evidence that children's invented spelling is a systemic representation of their sound-letter understanding (Chomsky, 1976). She showed similar patterns of invented spelling examples such as letter name strategy (e.g., spelling H for the /ch/ sound [BRENH for *branch*] and spelling EF for the /f/ sound [EFTR for *after*]) and nasal omissions before consonant (e.g., PASLS for *pencils*).



Figure 1
Developmental Stages of Spelling Knowledge

Building on these findings, a number of developmental stage models were proposed to describe qualitative changes in spelling knowledge across age and grade levels (Ehri, 1989; Gentry, 1981; Henderson & Templeton, 1986). Even though these models use different terms to classify children's spelling stages, the basic notions across the models are very similar. That is to say, all the models propose that children go through qualitatively different phases of language knowledge which are required for learning to spell: from alphabetic, to orthographic, and to morphemic understanding (Figure 1). Beginning learners of spelling need to know how sounds in words can be represented by letters. Letter names and sound of the letter

are essential for novice spellers. After that, it is necessary for children to understand what letter combinations are possible in the English writing system, namely orthographic knowledge. The final knowledge gained is an understanding of morphemic structure. With knowledge of inflectional, compounding, and derivational morphology, children can spell more complex, multisyllabic words as well as understand the multiple meanings they contain.

In response, strategy-based approaches emphasized that children draw on multiple sources of linguistic information simultaneously rather than sequentially. Rebecca Treiman's strategy approach does not strictly discriminate children's developmental trajectory of spelling skills. Treiman (1993) pointed out, is that children use various strategies to spell from beginning, rather than utilizing one skill at a time. This view has been supported by empirical evidence that children at the beginning stage have both orthographic and morphemic knowledge along with the ability to manipulate alphabetic strategies (Varnhagen et al., 1997). For example, first graders often use orthographic patterns in spelling to determine that *ck* does not occur at the beginning of English words (Treiman, 1993). In addition, first graders can spell the flap of *dirty* with a *t* rather than *d*, because they know *dirty* comes from *dirt* (Treiman et al., 1994). Based on this evidence, Treiman claimed that young children's spelling development is a function of the continual and concurrent accumulation of all alphabetic, orthographic, and morphemic knowledge, rather than process of stages (Treiman & Bourassa, 2000).

Although the strategy approach provides a more flexible account of spelling development, it still offers limited explanatory power regarding how different components of spelling knowledge become integrated within word-level representations. This issue is addressed more directly by the LQH (Perfetti, 2007). According to the LQH, spelling development reflects the increasing precision, stability, and integration of phonological, orthographic, and morphological information within lexical representations. High-quality lexical representations support efficient access and accurate production, whereas low-quality representations are underspecified, weakly integrated, or competing, resulting in spelling errors. A key implication of the LQH is that spelling accuracy alone is an insufficient indicator of development. Learners may show similar accuracy levels while relying on qualitatively different sources of information, and apparent plateaus in performance may mask ongoing lexical restructuring. From this perspective, spelling errors are not merely incorrect outcomes but developmental traces of partially specified lexical representations. Analyses that focus exclusively on overall scores or broad error descriptions therefore risk overlooking important developmental differences.

Feature-based assessment tools, such as the Primary Spelling Inventory (PSI), are particularly well suited to an LQH-oriented analysis of spelling development. By separating spelling performance into features such as initial and final consonants, vowel patterns, digraphs, consonant blending, and inflectional endings, the PSI allows researchers to examine which components of lexical knowledge are relatively stable and which remain fragile at different grade levels. When combined with error-type analysis, this approach enables a more fine-grained examination of how spelling development unfolds as a process of lexical quality growth rather than simple skill accumulation. In the context of Korean EFL learners, this perspective is especially important. Limited English input may constrain opportunities for implicit pattern learning, making the integration of phonological, orthographic, and morphological information more uneven across lexical features. Adopting a PSI-based, feature-level analysis grounded in the LQH, thus provides a theoretically motivated and methodologically sensitive framework for investigating spelling development in this population. The present study builds on this framework to examine grade-related changes in spelling accuracy, feature-level performance, and error distributions as indicators of lexical restructuring.

Korean EFL Learners' Spelling Development

There is a dearth of studies on the development of Korean EFL learners' spelling skills. Y. Lee (2007a, 2007b) pioneered studies that showed that Korean EFL learners' spelling skills follow similar developmental trajectories to those of native English speakers: from alphabetic knowledge, to orthographic, and finally to morphemic understanding. For example, fourth graders spelled the beginning sound of the word or consonants only, while fifth and sixth graders tried to spell orthographic and morphemic units of the target words. In particular, the study suggested that more proficient learners actively manipulate spelling patterns rather than relying solely on phoneme-grapheme correspondence. Interestingly, however, the study claimed that specific features of the Korean EFL learners' spelling development are not identical to their native English counterparts. Different from the explanation in the stage model, the Korean EFL learners acquire diphthongs (e.g., *play*, *mouth*, *eight*) earlier than they learn CVCC patterns (e.g., consonant-short vowel-consonant clusters; *help*, *want*, *sick*, *wash*). As such, Lee's work contributed to an early developmental understanding of English spelling in Korean elementary learners, moving beyond purely accuracy-based descriptions.

Nevertheless, the analytical framework proposed by Y. Lee (2007a, 2007b) exhibits theoretical limitations, particularly in its reliance on error complexity as the primary indicator of spelling development. While the sophistication scores

capture how differently learners misspell words, they do not specify which types of linguistic information—phonological, orthographic, or morphological—are integrated into lexical representations at different stages. For instance, features such as digraphs, vowel patterns, and inflectional endings are not analyzed as independent developmental components, nor is their stabilization traced systematically across grade levels. Consequently, developmental change is largely presented through increases in total scores or aggregate error sophistication, which implicitly frames spelling development as a linear and cumulative progression. This approach assumes that different aspects of spelling knowledge mature in tandem, an assumption that obscures potential asymmetries among subcomponents of lexical knowledge. As a result, Lee’s findings provide limited explanatory power regarding why certain spelling features remain fragile even as overall accuracy improves. From a contemporary perspective, this treatment risks oversimplifying the complexity of spelling development.

More recently, S. Lee (2010, 2019) sought to address Korean EFL elementary students’ English spelling through an analysis of invented spellings. According to S. Lee (2010), the frequency of spelling errors was gradually decreased from the fifth to sixth graders and sixth graders use more advanced linguistic knowledge (i.e., morphemic) than fifth graders (i.e., alphabetic). Based on the Gentry’s stage model, the researcher named the grade five and six Korean EFL learners are somewhere between phonetic and transitional stages, which refers to the alphabetic and the morphemic stages, respectively. S. Lee (2019) demonstrated that learners draw on partial phonological and orthographic knowledge even when their spellings are inaccurate. This perspective aligns with developmental theories that view spelling as a problem-solving activity rather than a simple reproduction task. However, S. Lee’s (2010, 2019) analysis primarily centered on individual cases and descriptive patterns, offering limited evidence regarding systematic developmental change across grade levels. Furthermore, the study did not quantify how frequently particular error types occurred, making it difficult to generalize the findings or link them to broader developmental trends.

Taken together, the studies by Y. Lee (2007a, 2007b) and S. Lee (2010, 2019) underscore the importance of viewing spelling errors as developmentally meaningful, yet they also reveal significant gaps that warrant further investigation. While prior studies primarily relied on overall accuracy or descriptive error patterns, they provide limited insight into how different linguistic components develop unevenly and interact within lexical representations. From the perspective of the LQH (Perfetti, 2007), spelling development should be understood as the gradual coordination and stabilization of multiple sources of linguistic information within lexical representations. To capture this process more precisely, a feature-based and distributional approach is required, allowing for the examination of both the strength of specific spelling features and the structure of error patterns. A feature-based approach, such as that afforded by the PSI allows researchers to disentangle which aspects of lexical knowledge are robust and which remain fragile at different grade levels. By combining accuracy, feature-level performance, and error-type analysis, the present study extends prior Korean EFL research by offering a more fine-grained account of spelling development as a process of lexical quality growth rather than simple skill accumulation.

The Present Study

The purpose of the present study is to investigate Korean elementary EFL learners’ spelling development from a LQH perspective, using a feature-based analysis of spelling performance. Specifically, this study aims to examine how spelling accuracy, feature-level performance, and error patterns vary across grade levels and what these patterns reveal about the quality of learners’ lexical representations. By employing the PSI, the study seeks to move beyond stage-based classifications and to capture fine-grained differences in the stability and integration of spelling-related knowledge. Through this analysis, the study aims to contribute to the literature by (a) providing empirical evidence on Korean EFL learners’ spelling development at the lexical feature level, (b) demonstrating the utility of the LQH as a theoretical framework for EFL spelling research, and (c) offering pedagogically relevant insights into areas of persistent difficulty and lexical fragility. The present study addresses the following research questions:

- 1) How does overall spelling accuracy vary across grade levels among Korean elementary EFL learners?
- 2) Which PSI spelling features demonstrate relative dominance within each grade level, and how does the dominance pattern changes across grades?
- 3) What types and distributional patterns of spelling errors are observed across grade levels?
- 4) What do these error patterns reveal about the quality and integration of learners’ lexical representations?

METHOD

Participants

A total of 139 elementary school students participated in the present study. As represents in Table 1, the sample consisted of 47 fourth graders (28 boys and 19 girls), 42 fifth graders (23 boys and 19 girls), and 50 sixth graders (26 boys and 24 girls), yielding an overall distribution of 77 boys and 62 girls. All participants were enrolled in public elementary schools in South Korea and were receiving regular English instruction as part of the national curriculum. With respect to English exposure, the vast majority of participants ($n = 131$, 94.2%) reported no experience of living in an English-speaking country or engaging in extended English learning abroad for more than one year. This indicates that the sample largely represented learners in a typical English as a foreign language (EFL) context with limited naturalistic exposure to English outside the classroom.

The study participants' self-reported English spelling study frequency (Table 1) showed that most students reported studying English spelling at least three times per week ($n = 110$, 79.10%), indicating a generally moderate to high level of engagement with spelling practice across grades ($N = 139$). In Grade 4 ($n = 47$), the largest proportion of students reported studying spelling everyday ($n = 21$, 44.7%), followed by those reporting three or four times per week ($n = 13$, 27.70%). Smaller proportions indicated studying once or twice per week ($n = 6$, 12.80%) or never ($n = 7$, 14.9%). In Grade 5 ($n = 42$), the most frequent responses were everyday ($n = 16$, 38.10%) and three or four times per week ($n = 15$, 35.70%). Fewer students reported studying once or twice per week ($n = 7$, 16.70%) or never ($n = 3$, 7.10%). In Grade 6 ($n = 50$), a majority of students reported studying spelling three or four times per week ($n = 26$, 52.00%) or everyday ($n = 19$, 38.00%). Only a small number of students indicated studying once or twice per week ($n = 3$, 6.00%) or never ($n = 2$, 4.00%).

In Table 1, students also reported the amount of time spent studying English spelling during each study session and the results indicate that the participants across grades ($N = 139$) engaged in English spelling study for at least 30 minutes per session ($n = 96$, 69.00%). In Grade 4 ($n = 47$), the most common study times were 30–60 minutes ($n = 15$, 31.9%) and more than 60 minutes ($n = 15$, 31.9%), while fewer students indicated less than 30 minutes ($n = 12$, 25.50%) and never ($n = 3$, 6.40%). In Grade 5 ($n = 42$), the majority of students reported studying either 30–60 minutes ($n = 16$, 38.1%) or more than 60 minutes ($n = 16$, 38.1%). Smaller proportions reported studying less than 30 minutes ($n = 5$, 11.90%) or never ($n = 3$, 7.10%). In Grade 6 ($n = 50$), similar patterns were observed. The largest proportions of students reported studying 30–60 minutes ($n = 17$, 34.0%) or more than 60 minutes ($n = 17$, 34.0%), followed by those studying less than 30 minutes ($n = 13$, 26.0%) and never ($n = 2$, 4.0%).

Table 1
Participant Characteristics and English Spelling Study Background (N = 139)

Categories		Gr. 4 ($n = 47$)	Gr. 5 ($n = 42$)	Gr. 6 ($n = 50$)
Gender	Boys	28	23	26
	Girls	19	19	24
Study Frequency	Everyday	21 (44.7%)	16 (38.1%)	19 (38.0%)
	3–4 times/week	13 (27.7%)	15 (35.7%)	26 (52.0%)
	1–2 times/week	6 (12.8%)	7 (16.7%)	3 (6.0%)
	Never	7 (14.9%)	3 (7.1%)	2 (4.0%)
Study Time	< 30 min	12 (25.5%)	5 (11.9%)	13 (26.0%)
	30–60 min	15 (31.9%)	16 (38.1%)	17 (34.0%)
	> 60 min	15 (31.9%)	16 (38.1%)	17 (34.0%)
	Never	3 (6.4%)	3 (7.1%)	2 (4.0%)

The PSI

The PSI was administered to evaluate the participants' spelling skill. The PSI is a qualitative test protocol to evaluate young children's spelling development (Bear et al., 2020). The inventory consists of a list of 26 words that begins with simple CVC words (e.g., *fan*, *pet*) and ends with inflectional endings (e.g., *chewed*, *wishes*). The columns from left to right represent the typical stages of spelling development: from the emergent (phonetic stage), to the letter name-alphabetic (alphabetic

stage), to the within word pattern (orthographic stage), and to the syllables and affixes (morphemic stage). Even though all the 26 words are recommended for late first, second, and third graders, the participants of the present study were EFL learners who have very limited access to the target sounds of the words in the PSI list. Thus, the revised PSI (see the Appendix) was developed to assess Korean EFL elementary students' spelling development in a manner that is both linguistically principled and instructionally aligned with the national curriculum.

The development process followed a systematic three-step procedure grounded in the structure and design principles of the original PSI, while adapting the instrument to the Korean EFL context. First, the overall structure of the original PSI word list was retained as the foundational framework for revision. Specifically, each target word was analyzed in terms of syllable count, consonant–vowel structure (e.g., CVC, CCVC, CVCC, CVCe), and the specific spelling features it was designed to assess. These features included initial and final consonants, short vowels, consonant digraphs, consonant blends, common vowel patterns, other vowel patterns, and inflectional endings. Second, the original PSI word list was examined in relation to the 2022 Revised English National Curriculum for Elementary Schools in South Korea. Words included in the official elementary core vocabulary list were retained without modification to ensure curricular relevance and ecological validity. For example, items such as *dig*, *hope*, *wait*, *dream*, *shouted*, and *riding* were preserved as they appear in both the original PSI and the national curriculum.

After that, for PSI items that were not included in the national core vocabulary list, replacement words were selected from the 2022 curriculum-based vocabulary pool. These substitute items were carefully chosen to match the original PSI words in terms of syllable count, phonological structure, and, critically, the target spelling features (e.g., initial and final consonants, short vowels, digraphs, blends, common and other vowel patterns, and inflected morphemes). For example, the original PSI item *blade* was replaced with *slide*, which preserves the same monosyllabic structure, consonant cluster onset (*bl-* vs. *sl-*), and long vowel pattern realized through the silent *e*. Similarly, *fright* was replaced with *bright*, as both words share a CCVC structure with an initial consonant blend (*fr-* vs. *br-*) and the vowel team *-igh*, representing a long /aɪ/ sound. The item *wishes* was replaced with *dishes*, as both are two-syllable words containing the consonant digraph *-sh-* and the inflectional morpheme *-es*, thereby targeting both orthographic and morphological knowledge. These examples illustrate that each substitution was not a simple lexical replacement but a principled alignment based on syllable structure, phonological patterns, and target spelling features. This substitution procedure ensured that the revised PSI continued to assess the same underlying spelling features as the original instrument while aligning all items with learners' instructional experiences.

Finally, to reduce cognitive load for elementary EFL learners and enhance the feasibility of administration, the total number of test items was reduced from 26 to 24, and the feature scoring scheme was refined. Specifically, each item was designed to assess a more focused set of target features rather than multiple overlapping features. For instance, in words such as *rich* or *cheap*, the primary target feature was the digraph *ch*. Therefore, previously assessed features such as initial consonants, final consonants, or vowel patterns were not redundantly scored within these items. This adjustment minimized task complexity and allowed for a more precise measurement of specific orthographic knowledge. In addition, the coding of *-ding* and *-tting* in the Appendix was intentionally designed to enhance scoring clarity and diagnostic precision. Specifically, *-ding* was used to determine whether learners correctly represented the inflectional ending *-ing* in orthographic form, whereas *-tting* was used to capture whether learners applied the consonant doubling rule appropriately before adding the *-ing* suffix. In this sense, the two codings were not intended to represent morphological units per se, but rather to differentiate between distinct orthographic processes underlying the surface forms. This approach allowed for more efficient and fine-grained scoring of learners' spelling performance by separating the assessment of inflectional morphology from the application of orthographic rules such as consonant doubling.

The final revised PSI consisted of 24 target words, each designed to assess one or more specific spelling features. Across the entire word list, feature coverage was balanced, with six items assessing initial consonants, six assessing final consonants, six assessing short vowels, six assessing consonant digraphs, six assessing consonant blends, six assessing common vowel patterns, six assessing other vowel patterns, and five assessing inflectional endings. This design allowed for both a total spelling score and feature-level scores to be calculated for each student. All words were spoken first, and then an example sentence of the given word was orally provided. After that, the target word was spoken one more time. Importantly, the revised PSI was designed not merely as a measure of spelling accuracy but as a diagnostic tool that captures qualitative aspects of learners' lexical representations. By coding students' spellings according to specific features, the instrument enables researchers to examine which components of lexical knowledge are well established and which remain fragile at different grade levels. This feature-based scoring system is particularly well suited for analyses grounded in the LQH, as it allows spelling development to be interpreted as a process of gradual coordination and stabilization of phonological, orthographic, and morphological information within word-level representations.

Procedures

Prior to data collection, two elementary school English teachers from schools located in Busan and Gwangju were recruited through professional recommendations. These teachers were invited to participate as expert reviewers in order to ensure the developmental appropriateness of the spelling assessment for elementary EFL learners. As part of this process, the teachers were asked to review the revised PSI word list in advance and to provide feedback on whether the target words were suitable for students at the elementary school level. Based on the teachers' feedback, several concerns regarding word appropriateness were identified. Specifically, words such as *shouted*, *tries*, and *dishes* were considered potentially challenging for elementary EFL learners due to their phonological and morphological complexity. In addition, the word *dig* was noted as being culturally unfamiliar to many Korean children, particularly those raised in urban environments, and therefore potentially problematic in terms of vocabulary knowledge rather than spelling ability. The word *fat* was also flagged as potentially sensitive, as it could evoke negative emotional responses or interpersonal discomfort among students.

The concerns raised during the expert review were carefully considered; however, all revised PSI word items, including *shouted*, *tries*, *dishes*, *dig*, and *fat*, were ultimately retained for the following two reasons. First, the primary purpose of the present study was not to establish the validity or reliability of the PSI as an assessment instrument, but rather to diagnose how participating students represent heard words in written English and to examine developmental patterns in spelling across grade levels. From this perspective, it was considered important to preserve the full range of target items, regardless of the concerns noted by experts, in order to capture a broader spectrum of learners' orthographic representations and potential difficulties. Second, it was assumed that the issues identified in the expert review might not necessarily align with students' actual spelling performance. As the feedback was based on teachers' subjective judgments, it was treated as a useful reference point rather than a definitive criterion for item selection. Therefore, the final decision was to retain all items in the revised PSI to allow for an empirical examination of learners' spelling behaviors, rather than preemptively excluding items based on perceived difficulty or appropriateness.

The PSI evaluation was administered following a standardized oral presentation procedure. For each target word, students first listened to a recorded pronunciation produced by a native English speaker. They then heard an example sentence containing the target word to clarify its meaning and contextual usage. Afterward, the target word was pronounced once more, and students were instructed to write the spelling of the word on a response sheet. Before the main assessment, two practice items were administered to familiarize students with the test format and procedure. After confirming that students understood the task, the spelling assessment was conducted in a single session, which took approximately ten minutes to complete.

Following data collection, students' spelling responses were scored in multiple stages. First, overall spelling accuracy was evaluated at the word level. Next, responses were scored using the PSI feature-based scoring system, with points assigned for specific orthographic features (e.g., consonants, vowel patterns, digraphs, and inflectional endings). Finally, grade-level differences in spelling performance were examined, and spelling error patterns were analyzed qualitatively to identify characteristic features of spelling development across grades. This multi-level analytic procedure allowed for a comprehensive examination of spelling development in terms of both accuracy and underlying lexical representation.

FINDINGS AND DISCUSSION

Spelling Development across Grades

Grades 4, 5, and 6 were examined with respect to overall spelling accuracy, PSI total scores, and each PSI feature score using descriptive statistics (Table 2). Spelling accuracy was operationalized as the total number of correctly spelled items, whereas PSI total scores reflected performance across the full range of features embedded in the revised PSI. Across all measures, mean scores increased monotonically from Grade 4 to Grade 6, indicating steady developmental growth in spelling performance across the upper elementary years. This pattern was evident not only for global indices such as spelling accuracy and PSI total scores, but also across individual PSI feature categories, including initial and final consonants, vowel patterns, digraphs, blends, and inflected endings. Notably, features associated with basic phonological–orthographic correspondences (e.g., consonants and short vowels) showed relatively higher mean scores and reduced variability by Grade 6, whereas features tapping morpho-orthographic knowledge, particularly inflected endings, exhibited lower means and greater variability across all grades. This distribution suggests that while foundational spelling features become increasingly stabilized with grade advancement, more complex morphological features remain an area of ongoing development even in

the later elementary years.

Table 2
Descriptive Statistics of Spelling Scores across Grades (N = 139)

Measure	PSI Target Features (Count)	Max.	Gr. 4	Gr. 5	Gr. 6
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Initial Consonants	<i>d (1), f (1), h (1), p (1), s (1), w (1)</i>	6	4.36 (1.85)	5.10 (1.51)	5.38 (0.90)
Final Consonants	<i>b (1), g (1), n (1), p (1), t (2)</i>	6	4.30 (1.73)	4.86 (1.47)	5.42 (0.97)
Short Vowels	<i>a (1), e (1), i (3), u (1)</i>	6	3.32 (1.81)	4.10 (1.75)	4.50 (1.34)
Digraphs	<i>ch (2), sh (2), th (2)</i>	6	2.23 (2.29)	3.60 (2.32)	4.78 (1.54)
Blends	<i>br (1), dr (1), gr (1), sl (1), st (1), tr (1)</i>	6	3.38 (2.14)	4.79 (1.91)	5.38 (1.18)
Common Vowels	<i>ai (1), ea (1), igh (1), a_e (1), i_e (1), o_e (1)</i>	6	1.81 (2.13)	3.38 (2.16)	4.26 (1.63)
Other Vowels	<i>aw (1), ew (1), oi (1), or (1), ou (1), ow (1)</i>	6	2.64 (2.10)	3.88 (2.06)	4.58 (1.44)
Inflected Endings	<i>-es (1), -ies (1), -ed (1), -ing (1), doubling -ing (1)</i>	5	0.72 (1.31)	1.62 (1.78)	2.58 (1.54)
	PSI Total	47	22.77 (12.73)	31.31 (12.99)	36.88 (8.33)
	Spelling Accuracy	24	7.57 (6.68)	12.19 (7.11)	15.86 (5.77)

Note. Max. = Maximum possible scores for each measure. The PSI target spelling features presented in the Appendix are arranged alphabetically within each column of this table. The number of participants was 47 in Grade 4, 42 in Grade 5, and 50 in Grade 6.

Table 3
ANOVA Results for Grade Differences of Spelling Scores

Outcome Measure	<i>F</i>	<i>p</i>	Pairwise Comparison	Mean Difference	Hedges' <i>g</i>	<i>p</i>
Spelling Accuracy	19.67	< .001	Gr. 5 — Gr. 4	4.62	0.66	< .001
			Gr. 6 — Gr. 5	3.67	0.57	.003
			Gr. 6 — Gr. 4	8.29	1.32	< .001
PSI Total	18.65	< .001	Gr. 5 — Gr. 4	8.54	0.66	< .001
			Gr. 6 — Gr. 5	5.57	0.52	.006
			Gr. 6 — Gr. 4	14.11	1.31	< .001

Note. *F* values are based on one-way analyses of variance with degrees of freedom (2, 136). Post hoc comparisons were conducted using Tukey's honestly significant difference (HSD) test. Mean differences are reported as the higher grade minus the lower grade.

To examine statistically significant differences among Grades 4, 5, and 6, one-way analyses of variance (ANOVAs) were conducted separately for overall spelling accuracy and PSI total scores (Table 3). For spelling accuracy, the ANOVA revealed a statistically significant main effect of grade, $F(2, 136) = 19.67, p < .001$, indicating reliable differences among the three grade levels. A parallel ANOVA conducted on PSI total scores likewise showed a significant grade effect, $F(2, 136) = 18.65, p < .001$. Tukey post hoc analyses demonstrated a consistent, ordered pattern across both outcome measures. For spelling accuracy, Grade 5 students scored significantly higher than Grade 4 students (mean difference = 4.62, $p < .001$), Grade 6 students outperformed Grade 5 students (mean difference = 3.67, $p = .003$), and the largest difference was observed between Grades 6 and 4 (mean difference = 8.29, $p < .001$). A comparable pattern emerged for PSI total scores, with significant differences between Grade 5 and Grade 4 (mean difference = 8.54, $p < .001$), between Grade 6 and Grade 5 (mean difference = 5.57, $p = .006$), and between Grade 6 and Grade 4 (mean difference = 14.11, $p < .001$).

Notably, analyses of both raw mean differences and standardized effect sizes (Hedges' *g*) revealed that spelling development follows a nonuniform trajectory, characterized by continued improvement accompanied by a deceleration in the magnitude of gains at higher grade levels. Specifically, the mean difference between Grades 5 and 6 was smaller than that between Grades 4 and 5 for both outcome measures: for spelling accuracy, the Grade 4–5 difference was 4.62 points, whereas the Grade 5–6 difference was 3.67 points, and for PSI total scores, the corresponding mean differences were 8.54 and 5.57 points, respectively. For spelling accuracy, the effect size for the Grade 4–5 comparison was moderate to large ($g = 0.66$) and exceeded that of the Grade 5–6 comparison ($g = 0.57$), whereas the cumulative difference between Grades 4 and 6 was large ($g = 1.32$). A parallel pattern was observed for PSI total scores, with effect sizes of $g = 0.66$ for the Grade 4–5 comparison, $g = 0.52$ for the Grade 5–6 comparison, and $g = 1.31$ for the Grade 4–6 comparison. Together, these patterns

imply that spelling development across upper elementary grades may involve a shift from rapid growth in overall accuracy and feature mastery toward more gradual refinement of existing orthographic and morphological representations.

Dominant Spelling Features across Grades

To examine whether specific PSI features exhibited grade-specific dominance and whether such dominance shifted across grades, feature-level accuracy rates were calculated within each grade and ranked to reflect relative dominance (Figure 2). By focusing on within-grade accuracy profiles rather than absolute comparisons across grades, this analysis aimed to identify which spelling features were most robustly established at each grade level and to evaluate whether patterns of feature dominance followed the sequential progression predicted by stage-based models of spelling development. Figure 2 shows consistently high accuracy for consonant-based features, substantial growth in orthographic features such as blends and digraphs across grades, and persistently lower accuracy for inflected endings despite steady improvement. Importantly, the visualized patterns further reveal that feature development is better characterized by relative dominance and shifting distributions rather than by a fixed, linear sequence of acquisition.

The Grade 4 profile revealed a clear concentration of dominance in consonant-based features. Initial consonants (72.70%) and final consonants (71.63%) showed the highest accuracy rates, substantially exceeding those of vowel-related and morpho-orthographic features. In contrast, digraphs (37.23%), common vowels (30.14%), and inflected endings (14.47%) exhibited markedly lower accuracy. This gradient suggests that spelling performance at this stage relies primarily on phonological representations, particularly consonantal information, while orthographic and morphological features remain weakly specified. At the same time, the presence of measurable accuracy across all feature types indicates that multiple sources of linguistic knowledge are already available, although they differ considerably in strength. Notably, blends (56.38%) slightly outperform short vowels (55.32%), indicating that consonant cluster processing is at least as robust as, and potentially more stable than, vowel representation at this stage. This pattern diverges from native developmental expectations, in which short vowels typically stabilize prior to more complex consonant structures, and instead suggests an early reliance on consonant-based phonological strategies.

The Grade 5 profile reflects a broad strengthening of both phonological and orthographic features. Consonant-based features remained highly accurate—initial consonants at 84.92% and final consonants at 80.95%—while orthographic features such as blends (79.76%) approached comparable levels of accuracy. Digraphs (59.92%) and other vowel-related features (64.68%) also demonstrated substantial gains. However, this pattern does not reflect a clear shift in feature dominance from consonantal to vocalic or orthographic features, as would be predicted by a strict stage-based model. Instead, multiple feature types appear to develop in parallel, suggesting that learners are increasingly coordinating phonological and orthographic information rather than transitioning from one dominant system to another. Crucially, vowel-related features do not exhibit a stable internal hierarchy: short vowels (68.25%), other vowels (64.68%), and common vowels (56.35%) cluster more closely together than expected under a linear developmental model, indicating that vowel knowledge is expanding in a distributed and non-sequential manner. Despite this overall growth, inflected endings (32.38%) remained substantially less accurate, indicating that morpho-orthographic knowledge had not yet been fully integrated into the spelling system.

The Grade 6 profile further reinforces this pattern of cumulative and overlapping development. Final consonants (90.33%), initial consonants (89.67%), and blends (89.67%) all reached similarly high levels of accuracy, reflecting a convergence of phonological and orthographic strength rather than a displacement of earlier-acquired features. Digraphs also showed strong performance (79.67%), confirming continued growth in orthographic complexity. In contrast, inflected endings, although improved (51.60%), remained the least accurate feature and did not emerge as a dominant component. This persistent lag is particularly noteworthy, as stage-based models would predict morpho-orthographic features to assume a central role at this stage of development. Furthermore, the vowel system undergoes a clear reorganization: 'other vowels' (76.33%) slightly surpass short vowels (75.00%), while common vowels (71.00%) remain closely aligned, demonstrating that vowel categories do not follow a fixed developmental order but instead reorganize dynamically over time.

Overall, the dominance patterns observed in Figure 2 do not appear to support a strict stage-based progression characterized by a sequential shift in feature dominance from phonological to orthographic and then to morphological features. Instead, spelling development in Korean EFL learners may suggest a strategic and integrative trajectory, in which multiple sources of linguistic knowledge are gradually strengthened and coordinated over time. Consonant-based features form an early and stable foundation that continues to be relied upon across grades, while orthographic features develop alongside them rather than replacing them. In particular, the consistently strong performance of blends relative to vowel features may indicate the sustained influence of consonant-based processing, while the non-linear reorganization of vowel

categories may reflect instability in orthographic mapping. Notably, morpho-orthographic features such as inflected endings remain comparatively underdeveloped even in upper elementary grades, which may suggest delayed integration of morphological knowledge. This pattern contrasts with developmental trajectories typically reported for native English learners, for whom morphological features become more robustly integrated at later stages (Ehri, 2000; Bear et al., 2020). The present findings therefore may suggest that spelling development in EFL contexts is not only slower but may reflect context-specific developmental patterns, potentially characterized by prolonged reliance on phonological strategies and incomplete consolidation of morphological representations.

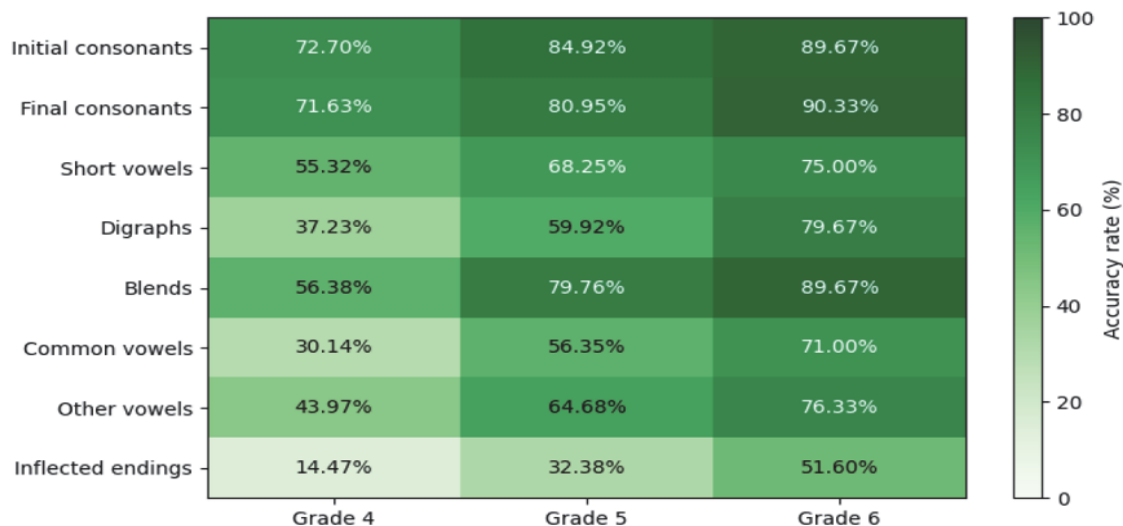


Figure 2

Feature-level Spelling Accuracy by Grade

Note. Values indicate accuracy rates (%). Shading represents accuracy level, with darker shading indicating higher accuracy.

Distributional Structure of Spelling Errors

As represented in Table 4, word-level error diversity was quantified for each PSI target word by aggregating all observed error forms and their frequencies across Grades 4–6 and computing two complementary indices: the type–token ratio (TTR) and normalized Shannon entropy (H_{norm}). TTR was calculated as the number of distinct error forms divided by the total number of error tokens for a given word and serves as a simple measure of how many different types of errors learners produced. In this sense, TTR reflects the range or variety of error forms. However, TTR does not account for how frequently each error form occurs. For example, a word may exhibit many different error types, but if most errors are concentrated in only one or two forms, TTR alone cannot capture this imbalance. To address this limitation, H_{norm} was additionally computed using Microsoft Excel. For each word, the relative frequency of each error form (p_i) was calculated by dividing its frequency by the total number of error tokens. Shannon entropy (H) was then calculated based on these relative frequencies so that both the number of different error forms and how evenly they were distributed could be taken into account (Shannon, 1948). To make fair comparisons across words with different numbers of error types, H was adjusted based on the number of distinct error forms (k), resulting in H_{norm} values ranging from 0 to 1. Unlike TTR, H_{norm} reflects not only how many different error forms occur but also how evenly errors are spread across them (Kuperman et al, 2021). When most errors are concentrated in one or two forms, H_{norm} is low, indicating that learners rely on a limited set of spellings. In contrast, when errors are spread more evenly across several forms, H_{norm} is high, indicating greater uncertainty and stronger competition among possible spellings.

Several words exhibited high error diversity, characterized by a large number of distinct error forms relative to error tokens and high H_{norm} values. For example, *down* produced 37 error tokens distributed across 31 distinct error forms, yielding a high TTR (.84) and the highest H_{norm} value (.98). Similarly, *new* (TTR = .72; H_{norm} = .97), *join* (TTR = .76; H_{norm} = .96), *think* (TTR = .70; H_{norm} = .95), and *bright* (TTR = .69; H_{norm} = .95) showed highly dispersed error distributions. These patterns indicate that errors for these words were spread relatively evenly across many alternative spellings rather than concentrated in a small set of dominant forms, suggesting ongoing competition among multiple partially specified orthographic representations.

In contrast, other PSI words showed lower diversity and more concentrated error distributions. The word *pen*, for instance, produced 46 error tokens but only 11 distinct error forms, resulting in a markedly low TTR (.24) and the lowest H_{norm} value (.58). This pattern indicates that a small number of recurrent misspellings accounted for most errors, consistent with a more constrained but still non-target like representation. Similarly, words such as *hope* (TTR = .42; H_{norm} = .72) and *grape* (TTR = .41; H_{norm} = .76) displayed moderate diversity, with errors clustering around a limited set of alternative spellings.

Importantly, high error volume did not necessarily correspond to high error diversity. For example, *shouted* produced the largest number of error tokens (88) and a large number of distinct error forms (52), yet its H_{norm} value (.83) was lower than that of words such as *down* or *bright*. This indicates that despite many observed error forms, the distribution of errors for *shouted* was relatively uneven, with certain misspellings occurring more frequently than others. Conversely, words with fewer total errors, such as *new* and *join*, showed very high H_{norm} , underscoring that diversity indices capture qualitative aspects of error distributions that are not reflected in raw error counts alone.

Taken together, these findings demonstrate that spelling errors vary not only in quantity but also in distributional structure across PSI words. Words with high H_{norm} values reflect less stabilized lexical representations characterized by multiple competing spellings, whereas words with lower H_{norm} values suggest partial convergence toward a smaller set of alternatives. From the perspective of lexical quality, this pattern indicates that spelling development involves differential stabilization across lexical items, with some words achieving more constrained representations earlier while others remain highly variable even in upper elementary grades.

Lexical Qualities in Spelling across Grades

When changes in TTR and H_{norm} were jointly examined across Grades 4–6, the PSI items exhibited multiple configurations of developmental change rather than a uniform trajectory (Table 4). The most prevalent configuration was Pattern 1 (TTR + - / H_{norm} + -), observed in 8 out of 24 items (33.3%), including *fat*, *bus*, *wait*, *slide*, *grape*, *cheap*, *draw*, and *sitting*. This pattern is best characterized as an inverted-U trajectory, in which both the diversity (TTR) and dispersion (H_{norm}) of error forms increase from Grade 4 to Grade 5 and subsequently decrease from Grade 5 to Grade 6. This pattern is clearly illustrated by *fat* and *grape*. For *fat*, TTR increased markedly from Grade 4 (.56) to Grade 5 (1.00), followed by a decrease in Grade 6 (.54). H_{norm} showed a similar trajectory, rising from .91 to 1.00 and then declining to .84. A comparable pattern was observed for *grape*, where TTR increased from .48 to .60 before decreasing to .41, and H_{norm} rose slightly from .86 to .87 before dropping to .71. These parallel shifts indicate that learners' spelling behavior becomes maximally variable and evenly distributed in Grade 5, before converging toward fewer and more dominant forms in Grade 6. Qualitative error patterns further support this interpretation. In Grade 4, learners produced a wide range of alternative spellings, reflecting early hypothesis generation; for example, *fat* elicited forms such as *pat*, *sad*, *fast*, and *pet*, as well as less phonologically aligned forms (e.g., *best*, *fet*), indicating unstable phoneme–grapheme mapping, while *grape* showed highly diverse responses including *great*, *graet*, *gray*, *greed*, and *grap*, along with analogical extensions such as *green* and *greek*, suggesting reliance on familiar rime patterns rather than target-specific representations. In Grade 5, variability reached its peak, with competing vowel representations in *fat* (e.g., *fate*, *fait*, *fit*) and diverse orthographic hypotheses in *grape* (e.g., *grade*, *grate*, *graph*), reflecting active but unstable rule application. By Grade 6, however, both the number of error types and their distribution became more constrained, indicating convergence; for *fat*, *pet* was the most frequent error, showing partial stabilization but persistent vowel confusion, while for *grape*, *great* emerged as the dominant error, indicating overgeneralization of a high-frequency orthographic pattern (-eat) and continued reliance on analogical processing rather than fully stabilized lexical representations.

Pattern 2 (TTR - - / H_{norm} - -), observed in 5 items (20.8%), including *pen*, *rich*, *dishes*, *shouted*, and *tries*, is characterized by a consistent decrease in both the number (TTR) and dispersion (H_{norm}) of error forms across grades. For example, *pen* shows a steady decline in TTR (.47 → .31 → .20) and H_{norm} (.77 → .76 → .66), while *dishes* (TTR: .79 → .68 → .50; H_{norm} : .97 → .89 → .76) and *shouted* (TTR: .85 → .67 → .57; H_{norm} : .96 → .88 → .82) exhibit similar downward trajectories. This pattern indicates that learners progressively reduce both the variety and distribution of error forms, suggesting convergence toward a smaller set of representations. However, qualitative analyses of the error forms reveal that this convergence does not necessarily reflect accurate target attainment, but rather a narrowing toward a limited set of persistent, non-target forms. In *pen*, Grade 4 responses include highly diverse and often unrelated forms such as *fan*, *pan*, *pet*, *ten*, and even non-phonological variants (e.g., *Dqa*, *qel*), indicating weak lexical specification. By Grade 5, the distribution becomes more restricted, with *fan* and *pan* emerging as dominant alternatives, and in Grade 6, responses are overwhelmingly concentrated on *pan*, with only a few residual variants (*fan*, *fen*). This suggests that learners stabilize on phonologically plausible but incorrect vowel representations rather than converging on the target form. In morphologically complex items such as *dishes* and *shouted*, the reduction in variability is accompanied by systematic simplification of

morphological structure. For *dishes*, Grade 4 responses include fragmented forms such as *dish*, *dis*, *dishas*, and *dishs*, suggesting incomplete encoding of both the base and the inflectional ending. By Grade 6, responses are heavily concentrated on *dish*, indicating stabilization on the base form with omission of the plural marker *-es*. Similarly, *shouted* shows extreme variability in Grade 4 (e.g., *shout*, *shout it*, *shoutd*), followed by increasing concentration on a limited set of forms such as *shout* and *show* in later grades. In Grade 6, *shout* dominates the distribution, suggesting failure to consistently encode the past tense marker *-ed*.

In contrast, Pattern 3 (TTR ++ / H_{norm} ++), observed in 3 items (12.5%), including *bright*, *down*, and *think*, is characterized by a monotonic increase in both the diversity (TTR) and dispersion (H_{norm}) of error forms across Grades 4–6. Unlike patterns that show eventual convergence, this configuration reflects sustained expansion without stabilization. For example, *bright* shows a steady increase in TTR (.50 → .83 → 1.00) and H_{norm} (.95 → .98 → 1.00), indicating that error forms become both more numerous and more evenly distributed over time. Similar trajectories are observed in *down* (TTR: .88 → .93 → 1.00; H_{norm} : .98 → .99 → 1.00) and *think* (TTR: .76 → .89 → 1.00; H_{norm} : .97 → .98 → 1.00), where both indices approach their maximum values by Grade 6. These patterns suggest that learners continue to generate a wide range of competing orthographic hypotheses, with no single form emerging as dominant. From a developmental perspective, this indicates prolonged instability in lexical representations, likely associated with phonologically complex structures such as consonant clusters (*think*) or diphthongs (*down*).

Table 4
TTR and Normalized Shannon H in Grades Four, Five and Six

PSI Items	All grades				Gr. 4				Gr. 5				Gr. 6				Across grades	
	Tokens	Types	TTR	H_{norm}	Tokens	Types	TTR	H_{norm}	Tokens	Types	TTR	H_{norm}	Tokens	Types	TTR	H_{norm}	Δ TTR	Δ H_{norm}
1. fat	42	24	.57	.88	18	10	.56	.91	11	11	1.00	1.00	13	7	.54	.84	+-	+-
2. pen	46	11	.24	.58	15	7	.47	.77	16	5	.31	.76	15	3	.20	.66	--	--
3. dig	48	26	.54	.91	12	11	.92	.99	23	16	.70	.92	13	10	.77	.93	-+	-+
4. bus	20	15	.75	.93	11	8	.73	.95	4	4	1.00	1.00	5	4	.80	.96	+-	+-
5. hope	45	19	.42	.72	25	10	.40	.71	13	5	.38	.79	7	7	1.00	1.00	-+	++
6. wait	33	17	.52	.90	17	11	.65	.94	11	9	.82	.98	5	3	.60	.96	+-	+-
7. stick	43	29	.67	.92	19	16	.84	.96	17	14	.82	.95	7	7	1.00	1.00	-+	-+
8. slide	50	36	.72	.93	19	15	.79	.96	17	17	1.00	1.00	14	9	.64	.89	+-	+-
9. dream	56	29	.52	.86	22	15	.68	.93	18	12	.67	.95	16	7	.44	.90	--	+-
10. grape	86	35	.41	.76	29	14	.48	.86	25	15	.60	.87	32	13	.41	.71	+-	+-
11. rich	77	39	.51	.90	24	18	.75	.96	23	17	.74	.94	30	15	.50	.93	--	--
12. cheap	73	33	.45	.87	15	7	.47	.77	31	17	.55	.91	27	14	.52	.89	+-	+-
13. bright	52	36	.69	.95	24	12	.50	.95	18	15	.83	.98	10	10	1.00	1.00	++	++
14. new	18	13	.72	.97	8	8	1.00	1.00	8	8	1.00	1.00	2	2	1.00	1.00	==	==
15. draw	55	37	.67	.91	16	13	.81	.98	18	18	1.00	1.00	21	12	.57	.86	+-	+-
16. dishes	63	34	.54	.82	19	15	.79	.97	22	15	.68	.89	22	11	.50	.76	--	--
17. north	51	37	.73	.90	18	16	.89	.98	21	17	.81	.97	12	10	.83	.98	-+	-+
18. shoutec	88	52	.59	.83	20	17	.85	.96	33	22	.67	.88	35	20	.57	.82	--	--
19. join	33	25	.76	.96	22	15	.75	.94	5	5	1.00	1.00	6	6	1.00	1.00	+=	+=
20. down	37	31	.84	.98	16	14	.88	.98	15	14	.93	.99	6	6	1.00	1.00	++	++
21. think	40	28	.70	.95	17	13	.76	.97	18	16	.89	.98	5	5	1.00	1.00	++	++
22. tries	72	31	.43	.83	29	18	.62	.92	20	12	.60	.88	23	12	.52	.87	--	--
23. sitting	74	29	.39	.78	25	10	.40	.86	25	16	.64	.88	24	10	.42	.80	+-	+-
24. riding	63	29	.46	.79	28	16	.57	.83	25	11	.44	.81	10	7	.70	.90	-+	-+

Note. TTR = Type Token Ratio; H_{norm} = normalized Shannon H; Δ TTR = TTR changes across grades 4-6; and Δ H_{norm} = normalized Shannon H changes across grades 4-6.

Qualitative analyses of the error forms further substantiate this interpretation by showing that the increase in both TTR and H_{norm} reflects not random variability, but the proliferation of structurally diverse yet competing orthographic hypotheses. In *bright*, Grade 4 responses already exhibit considerable variability, including forms such as *brit*, *brite*, *breat*, and *brot*, indicating unstable mapping of both the vowel and the consonant cluster. In Grade 5, this variability expands further, with forms such as *brait*, *bread*, *bret*, *brigt*, and *brite*, reflecting multiple competing attempts to encode the rime *-ight* using different vowel graphemes (*ai*, *ea*, *i*) and simplified codas. By Grade 6, rather than converging on the target form, learners continue to generate diverse alternatives such as *braed*, *braingt*, *briht*, *brite*, and *dright*, suggesting persistent instability in representing the vowel–consonant sequence /aɪt/ and the orthographic pattern *-ight*. The pattern is even more pronounced in *think*, where both onset and coda complexity contribute to sustained variability. In Grade 4, learners produce a wide range of substitutions for the interdental fricative /θ/ and the final cluster /ŋk/, including *thick*, *pink*, *thing*, *sink*, and *tink*. In Grade 5, variability increases further with forms such as *thing*, *thin*, *thinck*, *tik*, *tin*, and *ting*, indicating ongoing attempts to represent both the onset and the coda. By Grade 6, diverse forms such as *sing*, *thick*, *tingk*, and *tink* persist, showing that neither the onset nor the final cluster has been fully stabilized. Notably, the coexistence of phonologically plausible variants (e.g., *tink*, *tingk*) and more distant forms (e.g., *sing*) suggests that learners are simultaneously drawing on multiple representational strategies without resolving them into a unified lexical entry.

CONCLUSION

The present study investigated the developmental trajectories of spelling in Korean elementary EFL learners by examining not only overall accuracy but also the distributional properties of spelling errors through TTR and H_{norm} . By integrating these quantitative indices with feature-based analysis from the PSI, the study provides a multidimensional account of lexical development that extends beyond traditional accuracy-based approaches.

The findings demonstrate that spelling development in EFL learners is neither linear nor uniform across linguistic features. While overall spelling accuracy and PSI scores improved significantly across grades, the difference between Grades 5 and 6 was smaller than that between Grades 4 and 5, which may suggest a possible slowing of growth in the upper elementary grades. Feature-level analyses revealed that consonant-based features (e.g., initial and final consonants) were relatively well established even in lower grades, whereas orthographic patterns (e.g., digraphs, vowel patterns) and especially morphological features (e.g., inflected endings) remained persistently challenging. These results reflect the influence of English orthographic depth and the limited exposure typical of EFL contexts, where phonological transparency does not consistently map onto orthographic representation (Katz & Frost, 1992; Koda, 2005). Particularly in the Korean context, where learners' prior experience with a transparent writing system (Hangul) may lead them to rely heavily on phonological strategies when spelling English words. Such reliance may partially explain the persistent difficulty observed in vowel patterns, which require flexible grapheme selection, and in inflected endings, which involve morpho-orthographic transformations that are not directly recoverable from pronunciation. Furthermore, the relatively limited emphasis on systematic spelling instruction in elementary English education in Korea may contribute to delayed or incomplete integration of orthographic and morphological knowledge.

More importantly, the integration of TTR and H_{norm} analyses revealed three distinct developmental patterns that provide deeper insight into learners' lexical quality. Pattern 1 (TTR + – / H_{norm} + –) may be interpreted as reflecting successful lexical consolidation, in that initially diverse error forms are gradually reduced as learners appear to refine and stabilize their orthographic representations. Pattern 2 (TTR – – / H_{norm} – –) can be interpreted as suggesting constrained stabilization, in which learners produce fewer types of errors over time, but instead of becoming more accurate, they repeatedly rely on a small set of the same incorrect spellings. Pattern 3 (TTR + + / H_{norm} + +) may indicate prolonged instability, characterized by increasing diversity and dispersion of errors, which suggests that learners may continue to generate competing hypotheses without sufficient constraint.

These differentiated trajectories strongly support the claims of the LQH (Perfetti, 2007; Perfetti & Hart, 2002), which posits that high-quality lexical representations require the precise and integrated encoding of phonological, orthographic, and morphological information. The present findings suggest that EFL learners often operate with underspecified or weakly integrated representations, resulting in different developmental outcomes: successful consolidation following exploratory variability (Pattern 1), constrained stabilization around partially specified forms (Pattern 2), or prolonged competition among multiple alternatives (Pattern 3). Crucially, similar surface-level changes in variability (e.g., increases or decreases in TTR) may reflect fundamentally different underlying processes. This underscores the importance of jointly interpreting TTR and H_{norm} , rather than relying on a single index, to capture the dynamic nature of lexical development.

From a pedagogical perspective, these findings call for more explicit, systematic, and differentiated spelling instruction tailored to learners' developmental profiles. First, for Pattern 2 words, where learners exhibit reduced variability but incorrect stabilization, instructional emphasis should be placed on disrupting entrenched erroneous forms. This requires targeted corrective feedback, repeated retrieval practice, and error-based learning approaches. For example, teachers can implement “compare-and-contrast” activities where learners analyze their own misspellings against correct forms, or use delayed copying and self-correction tasks to promote deeper encoding. Providing immediate, specific feedback that highlights the mismatch between phonological expectation and orthographic convention is particularly critical (Graham et al., 2008; Hattie & Timperley, 2007). Additionally, incorporating spaced practice and cumulative review can help restructure fossilized representations. Second, for words exhibiting Pattern 3, where learners demonstrate high variability and ongoing competition among representations, instruction should focus on constraining hypothesis generation through explicit teaching of orthographic regularities. This includes structured phoneme–grapheme mapping activities (e.g., word building with letter tiles, dictation with immediate feedback), explicit instruction of vowel patterns and digraphs, and contrastive analysis of minimal pairs (e.g., *ship–sheep*, *bit–bite*) to refine orthographic distinctions. In addition, instructional support should explicitly target the resolution of competing representations by guiding learners toward dominant and conventional patterns, rather than merely increasing exposure to alternatives. Research has shown that such explicit and systematic instruction significantly improves spelling performance and transfers to reading outcomes (Galuschka et al., 2020; Graham & Santangelo, 2014).

In addition, given the persistent difficulty observed in morphological features across all grades, instruction should systematically incorporate morphological awareness training. This includes explicit teaching of inflectional endings (e.g., *-ed*, *-s*, *-ing*), morphological decomposition tasks (e.g., *jumped* → *jump* + *ed*), and word family analysis to highlight consistent spelling patterns across related forms. Prior research has demonstrated that morphological instruction enhances both spelling accuracy and vocabulary development, particularly in contexts where phonological cues are insufficient (Bowers & Kirby, 2010; Carlisle, 2010; Nunes et al., 1997). Importantly, these instructional approaches should not be applied uniformly but rather aligned with learners' observed error patterns. The present study suggests that TTR and H_{norm} can serve as diagnostic indicators for identifying whether learners are in a stage of exploration, consolidation, or constrained stabilization. However, their use in classroom settings should be considered as a complementary analytic tool rather than a direct assessment metric, given the practical constraints of real-time instructional contexts. Thus, incorporating distributional analysis into classroom assessment may enable more responsive and data-driven instruction, moving beyond traditional right–wrong scoring toward a more nuanced understanding of learners' underlying representations.

Several directions for future research emerge from this study. First, longitudinal studies are needed to examine whether the identified patterns represent stable developmental pathways or transitional phases. Second, experimental studies should investigate the effectiveness of pattern-specific instructional interventions, particularly for learners exhibiting Pattern 2 and Pattern 3 characteristics. Third, future research should explore how individual differences—such as phonological awareness, morphological awareness, and exposure to English input—interact with error distribution patterns. Finally, expanding the scope of analysis to include authentic writing tasks would provide greater ecological validity and allow researchers to examine how these patterns manifest in more complex literacy contexts. In addition, future studies should examine how instructional variables (e.g., explicit vs. implicit instruction, feedback type, and input frequency) mediate shifts between patterns over time.

Despite its contributions, this study has several limitations. First, the use of a cross-sectional design limits the ability to track individual developmental trajectories over time. Longitudinal studies are needed to confirm whether the identified patterns represent stable developmental pathways. Second, the PSI, while widely used, may not fully capture the range of orthographic and morphological complexity encountered in authentic writing contexts. Future research should incorporate more ecologically valid tasks, such as free writing or dictation with contextual variation. Finally, further investigation is needed to examine how instructional practices influence the emergence of different error patterns, particularly in relation to explicit versus implicit spelling instruction.

In conclusion, this study demonstrates that spelling development in EFL learners is best understood as a dynamic process involving the interaction of variability, competition, and consolidation in lexical representations. By combining feature-based and distributional analyses, the study provides a more nuanced framework for understanding spelling development and offers empirically grounded directions for both theory and instruction.

References

- Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2020). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (7th ed.). Pearson.
- Bowers, P. N., & Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing: An Interdisciplinary Journal*, 23(5), 515–537. <https://doi.org/10.1007/s11145-009-9172-z>
- Brown, A. (1990). A review of recent research on spelling. *Educational Psychology Review*, 2, 365–397. <https://doi.org/10.1007/BF01323630>
- Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement. *Reading Research Quarterly*, 45(4), 464–487. <https://doi.org/10.1598/RRQ.45.4.5>
- Chomsky, C. (1976). Approaching reading through invented spelling. *Theory and Practice of Early Reading*, 2, 43–65.
- Ehri, L. C. (1989). The development of spelling knowledge and its role in reading acquisition and reading disability. *Journal of Learning Disabilities*, 22, 356–365. <https://doi.org/10.1177/002221948902200606>
- Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*, 20(3), 19–36.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2020). Effectiveness of spelling interventions: A meta-analysis. *Educational Psychology Review*, 32, 437–462. <https://doi.org/10.1080/00461520.2019.1659794>
- Gentry, J. R. (1981). Learning to spell developmentally. *The Reading Teacher*, 34, 378–381.
- Graham, S., Harris, K. R., & Santangelo, T. (2008). Research-based writing practices and the common core. *Elementary School Journal*, 109(2), 170–182. <https://doi.org/10.1086/681964>
- Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing: An Interdisciplinary Journal*, 27, 1703–1743. <https://doi.org/10.1007/s11145-014-9517-0>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Henderson, E. H., & Templeton, S. (1986). A developmental perspective of formal spelling instruction through alphabet, pattern, and meaning. *The Elementary School Journal*, 86, 304–316. <https://doi.org/10.1086/461451>
- Katz, L., & Frost, R. (1992). The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Elsevier North-Holland.
- Koda, K. (2005). *Insights into second language reading*. Cambridge University Press.
- Kuperman, V., Bar-On, A., Bertram, R., Boshra, R., Deutsch, A., Kyröläinen, A.-J., Mathiopolou, B., Oralova, G., & Protopapas, A. (2021). Prevalence of spelling errors affects reading behavior across languages. *Journal of Experimental Psychology: General*, 150(10), 1974–1993. <https://doi.org/10.1037/xge0001038>
- Lee, Sun. (2010). A study on English spelling instruction based on the analysis of invented spellings of the Korean elementary students. *Foreign Language Education*, 17(1), 235–256.
- Lee, Sun. (2019). An investigation of Korean elementary students' knowledge on English spelling based on the analysis of their invented spelling. *Journal of Learner-Centered Curriculum and Instruction*, 19(4), 617–643.
- Lee, Yoon. (2007a). Aspects on English spelling of Korean elementary school students. *Foreign Languages Education*, 14(3), 231–256.
- Lee, Yoon. (2007b). Spelling accuracy and spelling sophistication of primary school children. *Primary English Education*, 13(3), 65–91.
- Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies. *Journal of Child Psychology and Psychiatry*, 38(3), 323–334. <https://doi.org/10.1037/0012-1649.33.4.637>
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins.
- Read, C. (1973). Children's judgments of phonetic similarities in relation to English spelling. *Language Learning*, 23, 17–38.
- Read, C. (1975). *Children's categorization of speech sounds in English*. National Council of Teachers of English.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. Oxford University Press.
- Treiman, R., & Bourassa, D. (2000). The development of spelling skill. *Topics in Language Disorders*, 20(3), 1–18.
- Treiman, R., Cassar, M., & Zukowski, A. (1994). What types of linguistic information do children use in spelling? The case of flaps. *Child Development*, 65, 1310–1329.
- Varnhagen, C. K., McCallum, M., & Burstow, M. (1997). Is children's spelling naturally stage-like? *Reading and Writing: An Interdisciplinary Journal*, 9, 451–481.

Appendix

The Revised Primary Spelling Inventory

No.	Original	Revised	PSI Features								Total	
			Initial	Final	Short	Dig.	Bl.	Common	Other	Inf.		
1	fan	fat	<i>f</i>	<i>t</i>	<i>a</i>							
2	pet	pen	<i>p</i>	<i>n</i>	<i>e</i>							
3	<i>dig</i>	<i>dig</i>	<i>d</i>	<i>g</i>	<i>i</i>							
4	rob	bus	<i>b</i>	<i>s</i>	<i>u</i>							
5	<i>hope</i>	<i>hope</i>	<i>h</i>	<i>p</i>				<i>o_e</i>				
6	<i>wait</i>	<i>wait</i>	<i>w</i>	<i>t</i>				<i>ai</i>				
7	gum	<i>stick</i>			<i>i</i>		<i>st</i>					
8	sled	slide					<i>sl</i>	<i>i_e</i>				
9	<i>stick</i>	<i>dream</i>					<i>dr</i>	<i>ea</i>				
10	shine	grape					<i>gr</i>	<i>a_e</i>				
11	<i>dream</i>	rich			<i>i</i>	<i>ch</i>						
12	blade	cheap				<i>ch</i>						
13	coach	bright					<i>br</i>	<i>igh</i>				
14	fright	new							<i>ew</i>			
15	chewed	draw							<i>aw</i>			
16	crawl	dishes				<i>sh</i>					<i>-es</i>	
17	wishes	north				<i>th</i>			<i>or</i>			
18	thorn	<i>shouted</i>				<i>sh</i>			<i>ou</i>		<i>-ed</i>	
19	<i>shouted</i>	join							<i>oi</i>			
20	spoil	down							<i>ow</i>			
21	growl	think				<i>th</i>						
22	third	tries					<i>tr</i>				<i>-ies</i>	
23	camped	sitting									<i>-ting</i>	
24	tries	<i>riding</i>									<i>-ding</i>	
25	clapping	.										
26	<i>riding</i>	.										
Total			6	6	6	6	6	6	6	6	5	47

Note. Initial = initial consonants; Final = final consonants; Short = short vowels; Dig. = digraphs; Bl. = blends; Common = common vowels; Other = other vowels; and Inf. = inflected endings. The “Original” column presents the original PSI word list, while the “Revised” column reflects the version used in the present study. Words shown in italics appear in both the original and revised versions.