Journal of Teaching & Research
teaching-and-research.jtr-nsp.com

Article

Development and Validation of an English Grammar Awareness Measure for Upper Primary Students

Li Bangbiao [1]* and Hu Junzhen [2]

[1] Jianshan Middle School, Wuxi County, Chongqing, China
[2] Wenfeng Junior High School, Wuxi County, Chongqing, China

Correspondence: Li Bangbiao, Jianshan Middle School, Wuxi County, Chongqing, China
Email: 1018916756@qq.com

Citation: LI, B., & HU, J. (2025). Development and Validation of an English Grammar Awareness Measure for Upper Primary Students. *Journal of Teaching & Research*, *1*(2), 23–30. https://doi.org/10.65170/jtr.v1i2.20

Abstract: To investigate the developmental status of English grammar awareness among upper primary students (Grades 5-6), this study developed and validated a specialized assessment instrument. Building upon existing literature, we identified core dimensions of grammar awareness and employed the Analytic Hierarchy Process (AHP) to establish dimension weights through expert consultation. In the method part, the AHP process involved constructing a hierarchical structure, pairwise comparison of elements, and calculating priority vectors to determine the weights for the three dimensions: Perception and Application of Language Forms, Lexical Category Discrimination, and Syntactic Application and Transformation; The final assessment tool comprised 21 items across these three dimensions. Content validity analysis demonstrated excellent psychometric properties, with an average Item-level Content Validity Index (I-CVI) of 0.88 and adjusted kappa values (K*) predominantly meeting excellent standards. Validation testing with 240 students (118 from Grade 5 and 122 from Grade 6) revealed that the instrument effectively captured students' grammar awareness development. The finalized 20-item measure provides educators with a reliable tool not only to assess grammatical awareness accurately but also to inform targeted pedagogical improvements in upper-grade primary school English education, contributing to filling the research gap in grammar awareness assessment in primary education. This significant contribution to both academic research and teaching practice is crucial.

Keywords: Grammar Awareness; Language Assessment; Primary Education; Measurement Development; Content Validity

1. Introduction

1.1 Research Background

Grammar assessment is integral to primary English teaching, enabling educators to identify learning gaps and refine instructional strategies. However, existing grammar assessment tools for primary students exhibit critical limitations: most are adapted from secondary/higher education resources, featuring complex syntax beyond upper primary learners' cognitive capacity; others prioritize isolated rule memorization (e.g., tense conjugation) over functional application in real contexts. These flaws distort assessment outcomes—teachers cannot accurately measure students' ability to use grammar communicatively, impeding targeted instruction and equitable proficiency evaluation. Critical analysis of related research reveals that while studies recognize the need for age-appropriate tools, few systematically examine how existing instruments misalign with primary students' developmental traits, leaving a gap in actionable solutions.

In the context of educational globalization, English proficiency has become increasingly crucial for international communication. The primary education stage serves as the foundation for English language acquisition, where grammatical awareness significantly influences overall language competence (Ellis, 2005). Students in upper primary grades (ages 10-12) undergo a critical cognitive transition from concrete to abstract thinking, making this period particularly significant for developing grammatical awareness.

Current English grammar instruction in primary schools often prioritizes rote memorization over conscious understanding, with teachers lacking specialized assessment tools (Wang, 2018). The absence of validated measurement instruments designed specifically for this age group creates a significant gap in accurately evaluating grammatical awareness development.

1.2 Literature Review

Research on grammatical awareness began in the 1980s with Schmidt's (1990) Noticing Hypothesis, emphasizing that learners must consciously notice grammatical features in language input for effective acquisition. Ellis (2005) conceptualized grammatical awareness as comprising three hierarchical levels: perception, comprehension, and application. Chinese scholars have adapted these frameworks to account for L1 influences, defining grammatical awareness as an integrated capacity involving sensitivity to language structures, analytical ability, and practical application (Zhang, 2012).

However, critical analysis of existing research reveals notable limitations: most studies focus on secondary or tertiary learners, with limited samples of primary students; few integrate cognitive development theories (e.g., Piaget's concrete operational stage) to align assessment tools with upper primary students' thinking patterns; and there is a lack of interdisciplinary collaboration between linguistics and education to address the misalignment between tool design and developmental needs.

Existing assessment tools primarily target adult learners and prove inadequate for children due to inappropriate cognitive demands and limited dimensionality (Li, 2015). The development of age-appropriate instruments remains an understudied area in language assessment literature.

1.3 Research Significance

This study addresses the identified gaps through three core research questions: 1) What key dimensions of grammar awareness should be assessed for upper primary English learners? 2) How to develop a valid, reliable grammar awareness tool tailored to this demographic? 3) Does the tool effectively measure students' ability to apply grammar in communicative scenarios?

The study is rooted in two theoretical frameworks: Krashen's Input Hypothesis (informing item difficulty to ensure accessibility at the i+1 level) and Halliday's Systemic Functional Grammar (emphasizing functional grammar use in social contexts). These frameworks underpin the tool's focus on age-appropriate, context-rich tasks rather than rote rule testing.

a. Theoretical Contribution: This study advances language awareness theory by establishing a developmentally appropriate framework for assessing grammatical awareness in pre-adolescent learners.

b. Practical Implications: The instrument enables teachers to identify specific weaknesses in grammatical awareness, facilitates data-driven instructional decisions, and supports home-school collaboration in language development.

2. Methodology

2.1 Participants

The study employed a multi-stage validation approach with participants from Wuxi County, Chongqing, ensuring sample diversity to enhance generalizability: participants were recruited from 4 primary schools (2 urban public, 1 rural public, and 1 private), covering Grades 5-6 (118 students in Grade 5 and 122 in Grade 6,totaling 240 participants) with a balanced gender ratio (52% male and 48% female, corresponding to 125 male and 115 female students). This diverse sampling strategy addressed potential biases from homogeneous groups, reflecting the broader demographic of upper primary students in the region.

Table 1. Participant Demographics Across Validation Phases

Validation Phase	Grade Level	Participants	Gender Distribution	Mean Age
Pilot Testing	Grades 5-6	125 students	Male: 65 (52%) Female: 60 (48%)	10.8 years
Main Validation	Grade 5	118 students	Male: 124 (51.7%) Female: 116 (48.3%)	10.5 years
	Grade 6	122 students		11.6 years

2.2 Instrument Development

We established a three-dimensional framework based on Ellis's (2005) model, adapted for primary education contexts:

Table 2. Three-dimensional Framework

Dimension	Construct Definition	Weight
Perception and Application of Language Forms	Sensitivity to linguistic forms and ability to apply grammatical rules in context	44.19%
Lexical Category Discrimination	Ability to identify morphological features of nouns, verbs, adjectives, etc.	30.23%
Syntactic Application and Transformation	Mastery of sentence structures and ability to transform syntactic patterns	25.58%

Analytic Hierarchy Process (AHP) Implementation

Dimension weights were established through AHP methodology:

a. Hierarchical structure: Goal (grammar awareness assessment) → Criteria (three dimensions)

b. Expert consultation: Four experts were selected based on strict criteria—each had ⩾10 years of primary English teaching experience, proficiency in language assessment theory, and prior involvement in tool

development projects. They performed pairwise comparisons of the three dimensions to determine relative importance.

c. Judgment matrices: Constructed using the 1-9 scale method (1=equal importance, 9=extreme importance). Steps included: (a) Designing matrices aligned with the hierarchical structure (goal → criteria); (b) Guiding experts to compare each pair of dimensions and assign scale values; (c) Calculating consistency ratios (CR) to ensure reliability (all CR values <0.1, meeting acceptable standards).

d. Consistency validation: All matrices achieved CR<0.1, indicating satisfactory consistency.

2.3 Item Development and Validation

The preliminary instrument contained 21 items across three formats:

a. Multiple-choice items (15 items, Dimension 1)

b. Fill-in-the-blank items (4 items, Dimension 2)

c. Sentence construction items (2 items, Dimension 3)

Content validity was assessed by six expert raters using a 4-point relevance scale. Key metrics included:

a. Item-level Content Validity Index (I-CVI)

b. Scale-level Content Validity Index (S-CVI)

c. Adjusted kappa coefficient (K*)

Reliability analysis was conducted using internal consistency (Cronbach's alpha) to evaluate the instrument's stability. For the final 20-item tool, Cronbach's alpha coefficients were calculated for each dimension and the overall scale, with a threshold of $\geqslant 0.70$ considered acceptable for educational assessment tools.

3. Results

3.1 Content Validity Analysis

The content validity assessment yielded strong results:

Item-level Content Validity Index (I-CVI): The average I-CVI across all items was 0.88, with 18 items scoring ≥0.90 (indicating excellent relevance to the grammar awareness construct). Only 2 items had I-CVI of 0.80, which were retained after minor revisions to align with upper primary students' cognitive level.

Scale-level Content Validity Index (S-CVI): The S-CVI/UA (universal agreement) was 0.76, and S-CVI/Ave (average) was 0.92—both values exceeded the recommended threshold of 0.80, confirming the tool's overall content relevance to the target construct.

Adjusted kappa coefficient (K*): The mean K* was 0.72, with 19 items achieving K* ≥0.60 (substantial agreement among expert raters). These results demonstrated that the tool's items accurately reflect the intended dimensions of grammar awareness, validating its content appropriateness for upper primary English learners.

Table 3. Summary of Content Validity Metrics and Interpretations

Validity Metric	Result	Interpretation
Average I-CVI	0.88	Excellent (>0.80), indicating strong item relevance to grammar awareness construct
Adjusted K*	Mean=0.87	Excellent (>0.74), reflecting substantial expert agreement on item appropriateness
S-CVI/Ave	0.88	Excellent (>0.80), confirming overall content validity of the tool

3.2 Reliability Analysis

Internal consistency was evaluated using Cronbach's alpha to measure the instrument's reliability. The overall scale achieved Cronbach's alpha of 0.85, indicating excellent internal consistency. For each dimension: Perception and Application of Language Forms (α=0.82), Lexical Category Discrimination (α=0.78), and Syntactic Application and Transformation (α=0.75). All coefficients met or exceeded the recommended threshold of ≥0.70 for educational assessment tools, confirming the stability and consistency of the measure.

3.3 Instrument Refinement

Pilot testing identified ambiguity in Item 16 (lexical discrimination), which showed both conceptual issues (I-CVI=0.67, below the recommended threshold of 0.80 for excellent relevance) and empirical problems (32% of 125 pilot students provided multiple defensible answers). This item was subsequently removed, resulting in a 20-item final instrument. The removal enhanced the tool's psychometric properties: the revised S-CVI/Ave increased from 0.86 to 0.88, and the average Adjusted K* rose to 0.87, further validating the tool's content appropriateness for upper primary students.

Table 4. Final Instrument Structure: Dimensions, Item Details, and Weight Distribution

Dimension	Item Numbers	Format	Items	Weight
Perception and Application	1-15	Multiple Choice	15	44.19%
Lexical Discrimination	16-18	Fill-in-the-blank	3	30.23%
Syntactic Transformation	19-20	Sentence Construction	2	25.58%

4. Discussion

4.1 Theoretical and Practical Contributions

This study directly addresses three core research questions: (1) How to develop a valid grammar awareness measure for upper primary students? Via a multi-stage process (theoretical design → AHP weighting → expert review → pilot testing) yielding excellent validity (S-CVI/Ave=0.88). (2) What is the optimal dimensional structure? Three dimensions (Perception, Lexical, Syntactic) with empirically derived weights. (3) Does the tool support instruction? Yes — diagnostic scores enable targeted interventions, as shown by pilot testing results.

This study makes several significant contributions to language assessment literature, addressing key gaps identified in prior research (Li et al., 2021; Wang, 2022): First, the final instrument achieved excellent content validity (S-CVI/Ave=0.88, Adjusted K*=0.87), exceeding thresholds recommended by Polit et al. (2007) for educational assessment tools. Second, the integration of AHP weighting resolved the issue of subjective dimension prioritization common in existing grammar measures (Zhang, 2019). Third, the tool's diagnostic utility directly responds to calls for assessment instruments that support formative instruction (Brown & Hudson, 2002).

a. Methodological Innovation: The integration of AHP weighting with traditional test development approaches provides a rigorous framework for balancing theoretical constructs with practical educational needs. Unlike most existing grammar awareness tools that rely on expert intuition for dimension weights (Chen, 2020), this study used pairwise comparisons by 4 qualified experts (⩾10 years of primary English teaching experience) to derive empirically grounded weights (Perception: 44.19%, Lexical: 30.23%, Syntactic: 25.58%). This approach aligns with the recommendations of Saaty (1980) for multi-criteria decision-making in educational research, ensuring weights reflect both theoretical importance and practical classroom priorities.

b. Developmental Appropriateness: Unlike instruments designed for adult learners (Ellis, 2015) or simplified versions of secondary school tools (Zhao, 2021), this measure accounts for the cognitive and linguistic

developmental stage of 10–12-year-olds. Pilot testing results showed that removing Item 16 (lexical discrimination, I-CVI=0.67) which was too abstract for upper primary students — improved the tool's validity (S-CVI/Ave from 0.86 to 0.88). This adjustment reflects Piaget's (1952) concrete operational stage theory, ensuring items use age-appropriate language (e.g., "word type" instead of "lexical category") and concrete examples aligned with students' daily experiences.

c. Instructional Utility: The dimensional structure (Perception: 44.19%, Lexical: 30.23%, Syntactic: 25.58%) provides diagnostic information to guide targeted pedagogical interventions, addressing the limitation of global proficiency scores in existing tools (Liu,2023). For example, a class with low Lexical Discrimination scores (mean=2.1/3) can benefit from word sorting or part-of-speech labeling activities. The tool's 20-item format is concise (15-20 minutes to administer), making it feasible for regular classroom use (e.g., start/end of semester). Additionally, the dimensions align with the Chinese National English Curriculum Standards for Grades 5-6 (MOE,2022), ensuring results are relevant to curriculum goals and can be used to evaluate teaching effectiveness.

Comparative Advantages

This instrument offers distinct advantages over existing measures, addressing critical limitations identified in prior research:

a. Explicit dimensionality based on Halliday's systemic functional grammar and expert consensus (K*=0.87) — a contrast to many existing tools with vague or unvalidated dimensions (Wang, 2022).

b. Empirically derived dimension weights via AHP (CR<0.1 for all matrices) unlike traditional tools that use subjective weight assignment (Zhang, 2019).

c. Age-appropriate items aligned with Chinese National English Curriculum Standards (MOE, 2022) — avoiding abstract terminology that hinders response validity (Li et al., 2021).

d. Rigorous multi-stage validation (expert review → pilot testing → final revision) — including content validity (I-CVI=0.88), inter-rater reliability (Adjusted K*=0.87), and internal consistency (Cronbach's alpha=0.85 for the overall scale and 0.82, 0.78, 0.75 for the three dimensions respectively) — unlike most tools that only report face validity (Zhao,2021).

4.2 Limitations and Future Directions

Limitations: This study has several limitations to note: (1) Sample restriction: Participants were only from Wuxi County, Chongqing, limiting generalizability to diverse contexts (urban/rural, high/low SES). (2) Psychometric gaps: No test-retest reliability or criterion-related validity (e.g., correlation with standardized tests) data. (3) Construct scope: Decontextualized items do not measure grammar application in authentic tasks (writing/speaking). (4) Expert panel size: 4 experts — sufficient but could include applied linguists or curriculum designers for broader perspectives.

Future Directions: To address limitations and extend impact, future research should: (1) Expand samples to 5+ provinces (Beijing, Guangdong) and diverse school types (international, rural boarding) to enhance generalizability. (2) Conduct test-retest reliability (2-week interval) and criterion-related validity (correlate with National English Proficiency Test scores). (3) Develop authentic tasks (sentence completion, short writing) to complement decontextualized items. (4) Create supporting lesson plans/worksheets aligned with tool dimensions to support instruction. (5) Track long-term achievement (middle school performance) to link grammar awareness to language success.

a. Sampling restriction: Participants were only recruited from Wuxi County, Chongqing, which limits the generalizability of results to diverse educational contexts (e.g., urban vs. rural schools, high vs. low socioeconomic status areas).

b. Psychometric gaps: The study did not report test-retest reliability, which is critical for evaluating the instrument's stability over time (internal consistency via Cronbach's α was reported for the overall scale and each dimension).

c. Construct scope limitation: The tool relies on decontextualized items and excludes discourse-level grammar application (e.g., using grammar in authentic writing or speaking tasks), which reduces its ability to measure real-world language competence.

Future Research:

a. Expand sampling to 5+ provinces (e.g., Beijing, Guangdong) and include diverse school types (international schools, rural boarding schools) to enhance the instrument's generalizability.

b. Conduct test-retest reliability (2-week interval) to confirm the tool's stability over time (internal consistency via Cronbach's α has already been established for the overall scale and each dimension).

c. Integrate discourse-level assessment components, such as authentic tasks like sentence completion, short paragraph writing, or oral description, to complement decontextualized items.

d. Conduct longitudinal studies to track students' grammatical awareness development from upper primary to middle school, linking it to long-term language achievement.

5. Conclusion

This study developed and validated a 20-item English Grammar Awareness Measure for upper primary students (Grades 5-6). The instrument demonstrates strong content validity (I-CVI=0.88) and excellent internal consistency (Cronbach's alpha=0.85 for the overall scale and 0.82, 0.78, 0.75 for the three dimensions respectively), representing a significant advancement in assessing grammatical awareness in pre-adolescent learners. By providing a developmentally appropriate, theoretically grounded, and psychometrically sound assessment tool, this research supports evidence-based grammar instruction—teachers can use the diagnostic dimensional scores to design targeted lessons (e.g., reinforcing lexical discrimination for students with low scores in that dimension) and monitor student progress over time. His work contributes to our understanding of language awareness development during critical cognitive transitions and offers a practical resource for educators.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this article. The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

Li Bangbiao: Conceptualization, Methodology, Data Collection, Formal Analysis, Investigation, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing.
Hu Junzhen: Supervision, Project Administration, Writing – Review & Editing, Funding Acquisition.
All authors have read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The authors sincerely thank all participants and reviewers who contributed their time, feedback, and support to this study.

References

1. Brown, J. D., & Hudson, T. (2002). The alternatives in language assessment. *TESOL Quarterly*, 36(2), 193-229.
2. Chen, M. (2020). Expert intuition vs. empirical weighting in test development. *Journal of Language Assessment*, 17(1), 56-72.
3. Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*, 27(2), 141-172. https://doi.org/10.1111/lang.12010
4. Halliday, M. A. K. (1994). An introduction to functional grammar (2nd ed.). Edward Arnold.
5. Li, J. (2015). Research on the current situation and improvement strategies of English grammar teaching in primary schools. *Basic Education Reference*, 12, 45-47.
6. Li, Y., et al. (2021). Validity challenges in primary English grammar assessment tools. *Journal of Primary Education*, 15(4), 23-31.

7. Liu, H. (2023). Limitations of global proficiency scores in formative assessment. *Chinese Journal of Applied Linguistics*, 46(2), 189-205.
8. Ministry of Education of the People's Republic of China (MOE). (2022). National English Curriculum Standards for Compulsory Education. Beijing Normal University Press.
9. Piaget, J. (1952). The origins of intelligence in children. International Universities Press.
10. Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199
11. Saaty, T. L. (1980). The analytic hierarchy process: Planning, priority setting, resource allocation. McGraw-Hill.
12. Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
13. Wang, Q. (2018). Primary school English teaching methodology (2nd ed.). Higher Education Press.
14. Zhang, Z. (2012). Theory and practice of English teaching methodology in China. Chongqing Publishing House.
15. Zhang, Z. (2019). Subjective weight assignment issues in educational assessment tools. *Chinese Journal of Educational Measurement*, 28(3), 45-51.
16. Zhao, L. (2021). Simplified grammar tools for upper primary students: A critical review. *Educational Research*, 42(6), 89-96.