

CFL-KG: A Knowledge Graph for Instructional Support in Chinese as a Foreign Language

Han Wang¹, Qianyu Wang¹, Yunshi Lan¹, Ye Wang¹, Yuanyuan Liang² and Anqi Ding²

¹School of Data Science and Engineering, East China Normal University, Shanghai, China

²International College of Chinese Studies, East China Normal University, Shanghai, China

Abstract

Second language acquisition (SLA) has increasingly incorporated AI-based support, such as intelligent tutoring systems, LLM-based writing assistants, and spoken-language coaching tools. However, many existing approaches remain weakly grounded in language curricula, proficiency standards, and explicit educational relations, which limits their ability to support standards-aligned teaching and assessment in a systematic way. This challenge is particularly salient in Chinese as a foreign language (CFL) and international Chinese language education, where teachers and researchers often need to work across multiple standards, heterogeneous language resources, assessment materials, learner-error evidence, and cultural content. In this paper, we present CFL-KG, a deployed knowledge graph for Chinese as a foreign language that organizes key resources for international Chinese language education within a multi-standard hierarchical ontology. Specifically, CFL-KG incorporates: (1) a layered ontology that organizes heterogeneous educational resources in CFL learning and teaching; (2) a large-scale knowledge base constructed from standards documents, textbooks, corpora, and expert-reviewed instructional resources; and (3) an end-to-end system that supports interactive retrieval, relation exploration, assessment traceability, and evidence-based educational workflows. To validate the quality and utility of CFL-KG, we conduct both intrinsic and extrinsic evaluations. The intrinsic evaluation examines coverage, structural quality, and the correctness of key educational relations. The extrinsic evaluation investigates whether the system improves representative teacher-facing tasks, including standards-aligned preparation, assessment traceability, and error-informed planning. Experimental results indicate that CFL-KG provides a usable and traceable knowledge infrastructure for international Chinese language education and supports complex professional tasks more effectively than document search and keyword-only retrieval.

Keywords

Educational Knowledge Graph, Standards Alignment, Assessment Traceability, International Chinese Language Education, Teacher-facing AI

1. Introduction

International Chinese language education increasingly depends on multiple parallel standards, including the *Chinese Proficiency Grading Standards for International Chinese Language Education* and successive HSK-related specifications [11, 12]. In realistic teaching and assessment workflows, however, relevant resources are rarely organized in one coherent structure. Teachers and researchers move manually across standard documents, vocabulary lists, grammar descriptions, question banks, cultural resources, learner corpora, and error records. This fragmentation does not only slow down retrieval; it also makes it difficult to answer evidence-sensitive questions such as why a question belongs to a knowledge point or why a learner error is relevant to a given instructional target.

These difficulties are especially pronounced in Chinese as a foreign language (CFL). Compared with many general course settings, CFL work involves a dense mixture of characters, words, grammar points, levels, cultural knowledge, and learner errors, often under multiple coexisting standards. As a result, many professional tasks are inherently relational rather than merely retrieval-oriented. A teacher preparing level-specific materials may need to move from level descriptors to language units, then to representative questions, common learner errors, and relevant cultural points. A researcher analyzing assessment coverage may need to trace a question backward to target knowledge points, level descriptors, and supporting explanations.

AI for Education Day, SIGKDD 2026, August 12, 2026, Jeju, Korea



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

From an educational perspective, this is not merely a systems-engineering problem. Constructive alignment emphasizes consistency among learning goals, instructional activities, and assessment [8]. Assessment validity and evidence-centered design require transparent links between assessment items and underlying knowledge constructs [9, 16, 17]. In second language acquisition, error analysis and interlanguage theory show that learner errors and their causes are pedagogically meaningful for intervention planning [18, 19, 20]. If the underlying knowledge structure is not computable and traceable, then higher-level AI support remains brittle for precisely those tasks that matter in education.

Existing educational knowledge graphs and AIED systems have shown that structured representations and intelligent tutoring tools can support resource organization, recommendation, and instructional support [6, 7, 1, 2, 4]. Yet many systems assume that the underlying knowledge organization already exists, or they operate in single-course and single-standard settings. In international Chinese education, this assumption is often invalid: **resources are heterogeneous, standards co-exist, and key relations such as item-knowledge and error-knowledge links still need to be explicitly built and validated.**

This paper addresses that gap through CFL-KG. We focus on teacher and researcher workflows in international Chinese education, including standards-aligned preparation, assessment tracing, and error-informed planning. In this setting, the graph serves as an explicit knowledge and evidence layer that supports professional educational tasks.

We focus on two research questions:

- **RQ1:** How can a multi-standard graph representation unify language knowledge, assessment resources, learner-error evidence, and cultural content for Chinese language education tasks?
- **RQ2:** To what extent does such a graph support representative teacher-facing tasks more effectively and more traceably than manual document search and keyword-only retrieval?

The contributions are threefold. First, we present a multi-standard knowledge representation for international Chinese education that organizes language units, assessment resources, error evidence, and cultural content in one three-layer schema. Second, we describe a deployed system that supports cross-standard retrieval, item traceability, error-linked evidence exploration, and corpus-backed lookup. Third, we report both intrinsic and extrinsic evidence to show graph quality and teacher-facing task value.

2. Related Work

Educational knowledge graphs have been used to organize instructional resources, model curricular structures, and support retrieval, recommendation, and precision teaching [1, 2, 3, 4, 5]. This line of work shows that graphs matter in education not only as storage backends but also as explicit representations of how educational concepts and resources connect. However, most existing educational KG studies focus on single-course or general educational settings [1, 2, 3, 4]. International Chinese education introduces additional modeling challenges: multiple coexisting standards, diverse language-unit types, and the need to explicitly validate relations among questions, learner errors, and knowledge points. The challenge is therefore not merely to build a graph, but to build a domain model that supports standards alignment, assessment traceability, and error-informed planning.

A second line of work concerns assessment interpretability and diagnosis. Research on knowledge tracing, cognitive diagnosis, and item-skill mapping emphasizes the need for clear relations between assessment items and underlying knowledge constructs [15, 16, 17]. Our work aligns with this tradition but operates at a more infrastructural layer: we first construct an inspectable relation layer between knowledge and items so that subsequent assessment analysis and AI support rest on explicit structure.

A third line concerns educational theory itself. Constructive alignment explains why objectives, instructional materials, and assessments should be coherently connected [8]. Evidence-centered design explains why educational judgments should rest on traceable evidence [9]. Cognitive load theory suggests that manual cross-resource and cross-standard information organization imposes additional information burden on teachers [10]. These theories do not dictate a graph schema directly, but they

explain why standards alignment, item tracing, and evidence packaging are educationally meaningful design choices.

Finally, graph-enhanced retrieval and GraphRAG-style work show that explicit relations can improve evidence organization in complex queries [21, 22, 23]. In educational settings, especially when the user needs to answer *why* rather than only *what*, relations themselves become more important than text similarity alone. CFL-KG differs by treating these relations as explicit first-class educational links and by evaluating them against teacher-support tasks.

3. KG Ontology and Deployed System Design

3.1. Design Goals

The system is designed around three practical goals derived from teacher and researcher workflows: standards-aligned preparation, assessment traceability, and error-informed planning. These goals correspond to educational needs associated with constructive alignment, evidence-centered assessment, and error analysis.

1. **Standards-aligned preparation.** Users should retrieve language and cultural resources by standard and level, and compare aligned content across standards.
2. **Assessment traceability.** Users should move from a knowledge point to related questions, or from a question back to target knowledge points, through inspectable evidence paths.
3. **Error-informed planning.** Users should connect knowledge points with learner errors, likely causes, and corpus-backed examples to support intervention design.

Taken together, these goals position the system as an infrastructure for professional educational work. The graph is designed to help users organize, trace, and verify educational evidence with lower effort across standards and resource types.

3.2. Three-Layer Schema

CFL-KG adopts a three-layer schema with explicit standards-alignment hubs. The top layer models language knowledge, the middle layer models assessment resources, and the bottom layer models cultural extensions; in addition, *Error* and *Level* nodes function as hubs for diagnosis and standards alignment.

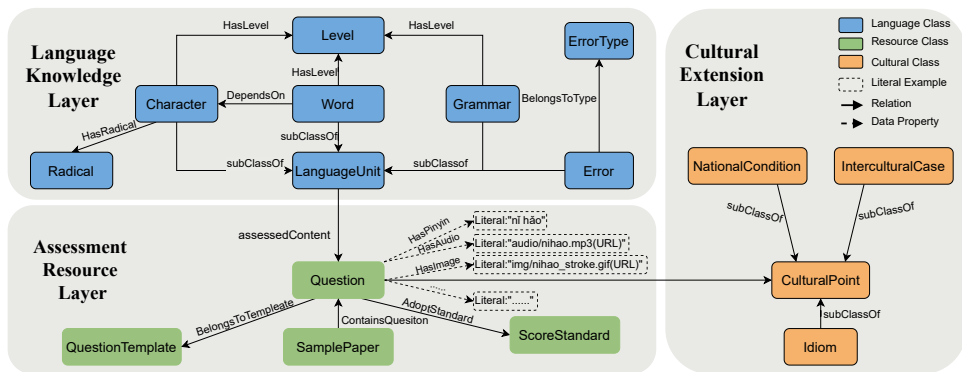


Figure 1: Schema of CFL-KG. The graph links language knowledge, assessment resources, learner errors, and cultural content through explicit educational relations and standards-alignment hubs.

The **language layer** centers on *Character*, *Word*, and *Grammar* entities linked through relations such as radical membership, composition, synonymy/antonymy, and learning dependencies. The **assessment layer** centers on *Question* nodes, which connect to target knowledge points and then to standards and template objects. The **cultural layer** contains *Cultural* and *Idiom* nodes that add

background knowledge and contextualized examples. *Error* nodes represent learner-error types, causes, and group labels and connect back to relevant knowledge points.

The educational importance of this representation lies not in merely aggregating resources, but in making task-relevant relations explicit. For example, item–knowledge links support coverage analysis and item interpretation; error–knowledge links support intervention planning; and level-alignment links support standards-based filtering and cross-standard comparison.

3.3. Deployed System Functions

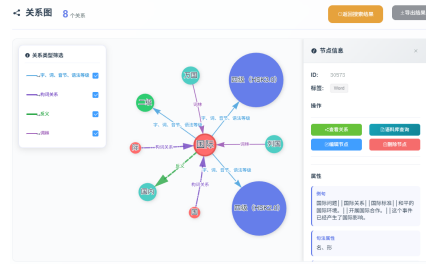
The deployed system stores graph data in Neo4j and exposes functions through a front-end interface and backend APIs. For this paper, four function families are most relevant: cross-standard retrieval, relation exploration, assessment traceability, and error-linked evidence exploration.

Cross-standard retrieval allows users to search nodes by type, keyword, and standard mode, then filter by level or entity class. Relation exploration supports one-hop and multi-hop expansion so that users can inspect links among knowledge points, questions, errors, and cultural nodes. Assessment traceability allows movement from knowledge points to related items and from items back to target knowledge points. Error-linked evidence exploration combines graph lookup with corpus context search so that users can pair structured error information with authentic examples.

Figure 2 presents representative query–result workflows in the deployed system. Panels (a)–(b) show browsing under different standard modes and inspection of the resulting relation graph for a selected entity. Panels (c)–(d) show Cypher or natural-language querying and the returned results in graph, table, and text views.



(a) Query: node type/scale overview.



(b) Result: relationship graph and node details.



(c) Query: Cypher/NL query workspace with templates.



(d) Result: visualization of retrieved subgraphs.

Figure 2: Representative query–result workflows in the deployed system. Top: browsing and relation inspection under standard modes; bottom: Cypher/natural-language querying and returned subgraph views.

4. KG Data Construction

CFL-KG is constructed from multi-source instructional and assessment resources under the graph schema in Section 3 [14]. The process combines deterministic parsing, rule-based normalization, statistical candidate generation, and expert review for high-impact labels, especially assessment-question–knowledge alignment, error annotation, and cross-standard alignment.

Sources and extraction. The graph integrates standards and examination documents, public language resources, and real-world teaching resources. Standards define levels, ability descriptors, and

standard-specific coverage [11, 12, 13]. Public dictionaries and vocabulary resources provide attributes for characters, words, radicals, pronunciation, grammar points, and examples. Teaching resources include question banks, reading materials, templates, corpus contexts, and learner-error logs. Structured tables were parsed with field mappings; semi-structured documents were segmented with type-specific templates. This separation is important because source resources differ in granularity: a standard entry may specify only a level and a language unit, whereas an assessment item contains a stem, options, answer, explanation, and sometimes implicit knowledge targets. The construction pipeline therefore preserves source provenance and intermediate evidence fields rather than only importing final nodes.

Normalization and linking. Each entity receives a stable identifier. Characters and words are normalized by simplified form, pinyin, radicals, and canonical labels; grammar points by name variants and example patterns; questions by stem, options, answer, explanation, type, and difficulty; and learner errors by error type, error form, likely cause, and supporting evidence. Deduplication combines exact matching, normalized-name matching, pinyin matching, and relation-context checks. Ambiguous cases are flagged for adjudication rather than automatically merged. For example, two grammar labels are merged only when their normalized names, examples, standard levels, and item links are consistent; otherwise, the graph keeps separate nodes with source and review status. This design reduces false merges that would otherwise contaminate assessment traceability.

Alignment and review. Cross-standard links are generated from source mappings and checked against level descriptors. Deterministic relations such as `FROM_LEVEL` and composition links are generated from standards and lexical structure; synonymy/antonymy is imported from curated resources; co-occurrence links are produced from corpus-window statistics. For `ASSESSES/EXAMPLE_OF` and `ERROR_RELATED_TO`, candidates are generated from keyword overlap, grammar labels, item explanations, and error descriptions, then reviewed by domain experts. Expert review focuses on whether a relation is educationally plausible, whether the source evidence supports the endpoint, and whether the relation direction is consistent with the schema. Approved, revised, pending, and rejected candidates are recorded so that later updates can be traced. The deployed graph contains over 20,000 characters, 250,000 words, 1,000 grammar points, 500 cultural nodes, 10,000 cloze/choice questions, and 2,000 reading materials.

5. Intrinsic Evaluation

The intrinsic evaluation asks whether the graph is reliable enough in coverage, structure, and key relations to support downstream educational tasks. We evaluate coverage/scale, schema and structural quality, and instance-level correctness of educational links. These dimensions correspond to practical educational usability: whether teachers can retrieve content by standards, whether relation paths remain interpretable, and whether high-value links are trustworthy enough for assessment tracing and intervention planning. We treat intrinsic evaluation as a data-quality audit rather than a model benchmark, because CFL-KG is intended to support evidence organization in instructional workflows.

Figure 3 characterizes the evaluated snapshot by node and relation types. The figure is used as dataset characterization rather than as an independent performance result.

Table 1
Intrinsic indicators for the evaluated CFL-KG snapshot.

Indicator	Value	Educational relevance
Entities aligned to at least one standard	97.8%	Standards-based filtering
Schema-violating edge rate	0.6%	Interpretable relation paths
Item–knowledge relation precision	92.0%	Assessment traceability
Error–knowledge relation precision	89.0%	Intervention planning
Level-alignment relation precision	95.0%	Cross-standard comparison
Overall reachability within ≤ 3 hops	91.2%	Short evidence chains



Figure 3: Type distribution of nodes and relations in CFL-KG.

5.1. Evaluation Protocol and Results

We evaluated three frozen construction snapshots: after schema locking, after relation-candidate construction, and after expert adjudication. Table 1 reports the final adjudicated snapshot, which is the version used in the deployed system and user study. Earlier snapshots were used to check whether later construction steps introduced schema violations, missing-standard links, or unexpected drops in short-hop reachability. Comparing snapshots was useful for identifying construction-stage errors: candidate generation increased graph coverage, while expert adjudication mainly improved relation precision and reduced unsupported links.

For instance-level correctness, we used stratified random samples of high-value educational relations: 220 item–knowledge links, 210 error–knowledge links, and 200 level-alignment links. Two raters with international Chinese education backgrounds judged endpoint type, relation direction, source evidence, and educational justification. Agreement was substantial (Cohen’s $\kappa = .84, .81,$ and $.90$ for the three relation groups). Disagreements were resolved through discussion with a third adjudicator. The raters used a simple rubric: a relation was marked correct only when both endpoints were appropriate, the direction matched the schema, and the evidence supported the pedagogical interpretation. Partial evidence or unclear endpoint typing was treated as incorrect for precision calculation.

The results indicate that CFL-KG has reached a usable level for the key relations required by representative teacher-facing tasks. Level-alignment precision affects the reliability of standards-aligned preparation; item–knowledge precision affects the credibility of assessment traceability; and error–knowledge precision affects whether intervention planning can rely on reasonable evidence. Short-hop reachability is also important: for teachers, graph value depends not only on theoretical connectivity, but on whether evidence chains can be inspected within a small number of operations.

6. Extrinsic Evaluation

6.1. User Study

The extrinsic evaluation tests whether the deployed system supports representative international Chinese education workflows better than manual document search and keyword-only retrieval. We used a within-subject design with 12 teachers or researchers. This design reduces individual differences and allows direct comparison among retrieval conditions. Participants had international Chinese education backgrounds; results were reported only in anonymized aggregate form. Before the formal tasks, participants completed a short system overview and warm-up exercise.

Each participant completed nine formal tasks: three task families under each of three retrieval conditions. Condition order was counterbalanced with a Latin-square schedule, and prompts were rotated so that no participant solved the same prompt twice. Prompts covered beginner and intermediate

CFL content and were matched by task type, standard, required evidence items, and difficulty. We avoided using the same surface keywords across conditions so that participants could not solve later prompts by memorizing earlier answers.

The three task families were: **T1** standards-aligned preparation, retrieving language/cultural points for a target standard and level and identifying relevant dependencies; **T2** assessment traceability, linking assessment items and target knowledge points or tracing from an item back to its knowledge target; and **T3** error-informed planning, identifying common learner errors, likely causes, and supporting practice or corpus examples. These tasks do not simulate an entire classroom process; instead, they isolate high-frequency steps that require cross-resource retrieval, relation verification, and evidence organization. For instance, a T2 prompt might ask participants to verify what a cloze item assesses and whether similar items overrepresent one grammar point; a T3 prompt might ask for likely causes and examples for a common learner error.

We compared manual search, keyword-only retrieval, and the full graph condition. Manual search allowed participants to use standards documents, item-bank materials, and conventional full-text or browser search. Keyword-only retrieval used the same underlying data and interface but disabled relation tracing and evidence-chain display. The full graph condition enabled relation tracing, multi-hop exploration, item–knowledge links, error links, and evidence views.

Outcome quality used a 3-point rubric aggregated to $[0, 1]$: 1.0 for correct and evidence-grounded answers, 0.5 for partially correct or insufficiently justified answers, and 0.0 for incorrect answers. This rubric fits semi-structured professional outputs where usability depends on both correctness and inspectable support. Two raters independently scored task quality and evidence completeness (Cohen’s $\kappa = .82$ and $.79$); disagreements were resolved before aggregation. We used Friedman tests and Holm-adjusted Wilcoxon signed-rank tests for within-subject comparisons.

6.2. Results

Table 2 summarizes the overall workflow outcomes. The full graph condition improves task quality and evidence completeness while reducing completion time. Keyword-only retrieval already reduces effort for simple lookup, but the full graph condition is more advantageous when users must verify mappings or organize multi-source evidence. Participants in the graph condition could follow explicit paths such as level–knowledge–item or error–knowledge–example instead of reconstructing these links manually from disconnected documents.

Table 2
Representative workflow outcomes across the three conditions.

Condition	Quality	Evidence	Median time (s)
Manual baseline	0.78	0.71	245
Keyword-only baseline	0.84	0.76	168
Full graph condition	0.91	0.88	109

Friedman tests indicated significant condition effects for task quality, evidence completeness, and completion time ($p < .05$). Holm-adjusted pairwise tests showed that the full graph condition significantly outperformed both baselines and was significantly faster. The advantage was clearest for T2 and T3, where users had to verify item–knowledge links and assemble intervention evidence across errors, causes, and examples. T1 showed a smaller gap because it contains more list filtering and less relation-intensive reasoning. These results suggest that the graph adds the most value when a task requires not just finding an object, but also explaining why resources are connected and which evidence supports the connection.

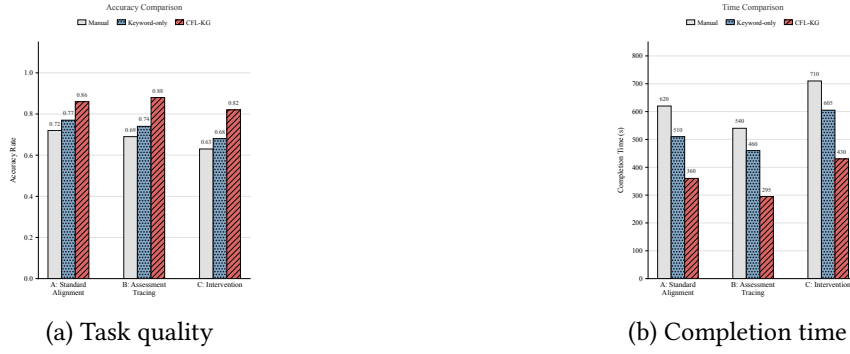


Figure 4: Task-based user study results across the three conditions.

7. Discussion

CFL-KG turns fragmented standards, knowledge points, assessment items, learner errors, and cultural resources into inspectable relation structures. Its contribution is not only faster retrieval, but support for traceable evidence chains in professional teacher and researcher tasks. This explains why gains are larger in T2 and T3, where users must verify mappings and combine evidence across resource types. The results also show a practical boundary: for simple lookup tasks, keyword retrieval may be sufficient, but graph structure becomes important when users need alignment, diagnosis, and justification.

CFL-KG supports teachers and researchers rather than autonomous high-stakes decisions about learners. Learner-error logs were anonymized before graph construction, and no personally identifying information is exposed in the deployed interface or paper. Question banks, reading materials, and learner-error evidence are controlled according to source permissions; rights-restricted teaching materials are used as internal evidence objects rather than redistributed as open data. The study is limited by its modest sample and controlled tasks, and future work will extend evaluation to authentic teaching settings and improve item-knowledge and error-related coverage. Another limitation is that some cultural and learner-error links still depend on expert judgment; future releases will expose confidence and review-status metadata so that teachers can distinguish reviewed evidence from automatically generated candidates.

8. Conclusion

We presented CFL-KG, a multi-standard knowledge graph for teacher-facing Chinese language education tasks. It unifies language knowledge, assessment resources, learner errors, and cultural content under an explicit schema and supports standards-aligned preparation, assessment traceability, and error-informed planning. Intrinsic and extrinsic evaluations show that CFL-KG provides a domain-specific evidence infrastructure for AI-supported international Chinese education. The camera-ready revision further clarifies the construction protocol, annotation procedure, task design, statistical analysis, and data governance boundaries, making the contribution more reproducible and easier to interpret. More broadly, the results suggest that a knowledge graph for CFL instruction should be evaluated not only by graph size, but by whether teachers can use its relations to justify instructional choices. This is why CFL-KG emphasizes inspectable standards links, item-knowledge links, error-knowledge links, and cultural extensions. These components make the dataset usable as a reusable evidence layer for future teacher-support tools and retrieval-augmented educational agents.

Declaration on Generative AI

The authors used generative AI tools for language polishing, grammar checking, and assistance in preparing the camera-ready revision. The research design, data analysis, claims, figures, tables, and final content were reviewed, verified, and approved by the authors.

References

- [1] Chen, P., Lu, Y., Zheng, V.W., Chen, X., Yang, B.: KnowEdu: a system to construct knowledge graph for education. *IEEE Access* **6**, 31553–31563 (2018)
- [2] Dang, F.-R., Tang, J.-T., Pang, K.-Y., Wang, T., Li, S.-S., Li, X.: Constructing an educational knowledge graph with concepts linked to Wikipedia. *Journal of Computer Science and Technology* **36**(5), 1200–1211 (2021)
- [3] Li, Y., Liang, Y., Yang, R., Qiu, J., Zhang, C., Zhang, X.: CourseKG: an educational knowledge graph based on course information for precision teaching. *Applied Sciences* **14**(7), 2710 (2024)
- [4] Qu, K., Li, K.C., Wong, B.T.M., Wu, M.M.F., Liu, M.: A comprehensive survey of knowledge graphs in education. *Electronics* **13**(13), 2537 (2024)
- [5] Abu-Salih, Z., Alotaibi, M.: Knowledge graph-based recommendation systems in education: a systematic literature review. *Heliyon* **10**(1), e23851 (2024)
- [6] Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A.: Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* **8**, 30–43 (1997)
- [7] Woolf, B.: *Building Intelligent Interactive Tutors*. Morgan Kaufmann, Burlington (2009)
- [8] Biggs, J.: Enhancing teaching through constructive alignment. *Higher Education* **32**(3), 347–364 (1996)
- [9] Mislevy, R., Steinberg, L., Almond, R.: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives* **1**(1), 3–67 (2003)
- [10] Sweller, J.: Cognitive load during problem solving: effects on learning. *Cognitive Science* **12**(2), 257–275 (1988)
- [11] Ministry of Education of China, National Language Commission: *Chinese Proficiency Grading Standards for International Chinese Language Education* (2021)
- [12] Chinese Testing International: *HSK-related specifications and exam framework documents* (2010–2023)
- [13] Chinese Testing International (Hanban) / related authorities. *HSK (New) / HSK 2.0 / HSK 3.0-related specifications and exam framework documents*. (Accessed via official releases and public materials), 2010–2023.
- [14] M. Hogan, E. Blomqvist, M. Cochez, et al. Knowledge Graphs. *ACM Computing Surveys*, 54(4):71, 2021.
- [15] Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* **4**(4), 253–278 (1995)
- [16] Tatsuoaka, K.: Rule space: an approach for dealing with misconceptions based on item response patterns. *Journal of Educational Measurement* **20**(4), 345–354 (1983)
- [17] De la Torre, J.: The generalized DINA model framework. *Psychometrika* **76**(2), 179–199 (2011)
- [18] Corder, S.P.: The significance of learners’ errors. *International Review of Applied Linguistics* **5**(4), 161–170 (1967)
- [19] Selinker, L.: Interlanguage. *International Review of Applied Linguistics* **10**(3), 209–231 (1972)
- [20] Ellis, R.: *The Study of Second Language Acquisition*. Oxford University Press, Oxford (1994)
- [21] Lewis, P., Perez, E., Piktus, A., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. In: *NeurIPS* (2020)
- [22] Zhu, X., Xie, Y., Liu, Y., Li, Y., Hu, W.: Knowledge graph-guided retrieval augmented generation. In: *NAACL* (2025)

- [23] Knollmeyer, K., Caymazer, O., Grossmann, D.: Document GraphRAG: knowledge graph enhanced retrieval augmented generation for document QA. *Electronics* **14**(11), 2102 (2025)