

Construction of an English Disciplinary Literacy Gradient Analysis Scale Based on the New Gaokao

Liu Yumei^{1,a,*}, Huang Yuanyan^{2,b}, Xie Jialuo^{3,c}

¹School of Foreign Languages, Lingnan Normal University, Zhanjiang, China

²Lian Jiang City Experimental School, Zhanjiang, China

³Zhanjiang Shaolin School, Zhanjiang, China

^a1045944776@qq.com, ^b19126697306@163.com, ^c2140889871@qq.com

* Corresponding author

Abstract: The English Literacy Gradient Analysis Scale under the New Gaokao (China's National College Entrance Examination Reform) framework is theoretically grounded in the English Curriculum Standards for Senior High Schools (2017 Edition, 2020 Revision) and the China Standards of English Language Ability, incorporating the Flesch Reading Ease formula to construct a five-dimensional framework: Discourse Form (textual function, discursive scope, rhetorical architecture) reveals the graduated design across informational transmission (A1), literary-aesthetic appreciation (A2), and specialized academic contexts (A3), mapping textual functionality onto cognitive hierarchies; Background Knowledge (situational knowledge D1, cultural knowledge D2, disciplinary knowledge D3) quantifies cross-disciplinary transfer demands, foregrounding Gaokao's assessment of cultural interpretation and epistemic integration; Linguistic Knowledge (knowledge density E1-E3, grammatical/lexical strata F1-F3, lexical breadth G1-G3) anchors the deep integration of contextual and rhetorical meanings; Knowledge Mobilization Pathways (H1 direct retrieval to H3 evaluative reconstruction) operationalize Bloom's higher-order thinking specifications, foregrounding critical deconstruction and creative knowledge application; and Surface Structure Processing Load (I1-I3) triangulates algorithmic metrics with qualitative analysis. The scale discloses five transformative imperatives for high school English teaching, learning, and assessment: (1) discourse instruction must adopt thematic text clusters, strengthening structural parsing and functional literacy; (2) knowledge scope should dismantle disciplinary silos, activating situational and cultural knowledge transfer to address authentic problems; (3) language skills require multimodal integration through unified viewing-listening-speaking-reading-writing-translation tasks to enhance evaluative synthesis; (4) disciplinary connotation must shift from formal meaning to rhetorical meaning, using context to drive critical consciousness; and (5) educational value requires cross-cultural dimensions, cultivating cultural rationality and global competence. This scale provides a quantifiable, operational instrument for systematic, integrated assessment of core competencies, empowering teachers to precision-design tiered instruction and facilitating competency-based educational transformation.

Keywords: New Gaokao; English Disciplinary Literacy Gradient Analysis Scale; Integration of Quantitative and Qualitative Analysis; Core Competencies

1. Introduction

The deepening of China's New College Entrance Examination (Gaokao) reform has positioned the scientific rigor and nuance of English language assessment as a critical determinant of pedagogical effectiveness and core competency cultivation. Prevailing evaluation instruments remain fragmented, targeting discrete linguistic skills or knowledge points without systematically integrating key literacy dimensions—discourse context, intercultural awareness, and cognitive quality. This limited scope yields insufficient washback for instructional improvement. A comprehensive, literacy-gradient analytic scale is therefore essential to structure English assessment within the reformed Gaokao framework. Synthesizing the *Senior High School English Curriculum Standards (2017 Edition, 2020 Revision)* (hereinafter referred to as the *New Curriculum Standards*), the *China's Standards of English Language Ability*, and the Flesch Reading Ease formula, this scale comprises five dimensions—discourse form, background knowledge, linguistic knowledge, answer retrieval complexity, and text readability—operationalized through 27 indicators across a three-tiered coding scheme (Levels 1–3). This architecture explicates the underlying design principles of Gaokao items, mapping their knowledge coverage, ability differentiation, and aesthetic value orientation. Through

systematic analysis of authentic test papers, the scale deciphers test developers' focal emphases, thereby informing pedagogical reconceptualization, curricular content selection, and evaluation methodology optimization, ultimately advancing teachers' assessment literacy and item construction competence.

Dimension construction is anchored in the academic quality level specifications of the new curriculum standards, foregrounding the alignment of discourse functions—including informational transfer, literary appreciation, and specialized academic contexts—with hierarchical cognitive demands. This mapping resonates with higher-order skill cultivation in Bloom's Taxonomy. For example, the discourse form dimension (encompassing text type, length, and rhetorical structure) quantifies not merely physical text features but also correlates with test-takers' linguistic intuition and capacity for interdisciplinary knowledge synthesis, thereby catalyzing a pedagogical shift toward thematically clustered texts and multimodal resource integration. The background knowledge dimension (everyday, cultural, and disciplinary knowledge) foregrounds the transcultural orientation of English education, mandating that instruction transcend superficial linguistic codes to engage with sociocultural dynamics and comparative analysis, thus fostering global perspectives and critical consciousness. The linguistic knowledge dimension (knowledge density, grammatical/lexical sophistication, and vocabulary breadth) strengthens the integration of contextual and rhetorical meaning-making, aligning with the curriculum's "form-meaning-function" triadic requirement and steering instruction from mechanical memorization toward communicative praxis.

This study's principal contribution resides in its hybridization of quantitative metrics (Flesch formula) with qualitative analysis, yielding actionable assessment criteria that narrow the divide between high-stakes test design and classroom practice. By deconstructing the graduated architecture of answer retrieval complexity—from direct identification to multi-source inferencing and comprehensive evaluation—the scale equips educators to scaffold tiered instructional activities, thereby fostering genuine "teaching-learning-assessment" integration. Subsequent sections will anatomize each dimension's competency connotations and derived pedagogical imperatives for senior high school English, offering empirical substantiation for systemic reform, catalyzing teacher professional growth, and ultimately serving the fundamental mandate of holistic, character-centered education.

2. The Literacy Connotations of Scale Dimensions and Their Indicators

Operationalized across five dimensions through twenty-seven granular indicators, this scale articulates a comprehensive framework for analyzing the New Gaokao across linguistic, intercultural, and cognitive planes. Elucidating the competency conscriptions embedded within each dimension reveals the hierarchical literacy standards that differentiate performance levels and operationalize core construct specifications. Taking the part of Multiple-Choice Reading and Banked-Filling of the New Gaokao as example, the complete analytic architecture is presented in Table 1.

Table 1: The Analysis Scale of New College Entrance English Examination

Analytical Dimension			Sample Codes																								
			Multiple-Choice Reading																			Banked Gap-Filling					
			PA	QA1	QA2	QA3	PB	QB1	QB2	QB3	PC	QC1	QC2	QC3	QC4	PD	QD1	QD2	QD3	QD4	PE	QE1	QE2	QE3	QE4	QE5	
Discourse Form	Functional Text Type	Informational Transmission (A1)																									
		Literary Aesthetics(A2)																									
		Specialized Academic(A3)																									
	Discursive Scope	Total Passage Word Count (B1)																									
		Total Stem Word Count (B2)																									
		Total Test Word Count (B3)																									
	Rhetorical Architecture	Canonical Architecture (C1)																									
		Latent Architecture (C2)																									
		Fragmented Architecture (C3)																									
Background Knowledge	Situational Knowledge (D1)																										
	Cultural Knowledge (D2)																										
	Disciplinary Knowledge (D3)																										
Linguistic Knowledge Architecture	Knowledge Density	Single-Point Activation (E1)																									
		Dual-Point Integration (E2)																									
		Multi-Point Synthesis (E3)																									
	Semantic Strata	Formal (F1)																									
		Contextual (F2)																									
		Rhetorical (F3)																									
	Lexical	Core																									

	Breadth	Vocabulary (G1)																								
		Extended Vocabulary (G2)																								
		Specialized Vocabulary (G3)																								
Knowledge Mobilization Pathways	Direct Retrieval (H1)																									
	Synthesized Inference (H2)																									
	Evaluative Reconstruction (H3)																									
Text Readability	Low Processing Load (I1)																									
	Moderate Processing Load (I2)																									
	High Processing Load (I3)																									

2.1 Discourse Form

This dimension operationalizes textual variation across three sub-constructs: functional text type, discursive scope, and rhetorical architecture.

2.1.1 Functional Text Type

Classification proceeds not from conventional genre categories (narration, exposition, argumentation), but from textually-actualized communicative functions evident in authorized curricula and recent Gaokao administrations. This reconceptualization is theoretically necessary: traditional genres resist linear difficulty scaling—expository complexity may exceed argumentative sophistication, while narrative accessibility varies widely. More critically, Chinese learners transfer baseline generic awareness from L1 instruction, yet demonstrate profound deficits in English textual function recognition and appreciative capacity. This functional illiteracy obstructs strategic information extraction, authorial intent synthesis, and aesthetic-critical evaluation, constricting expressive versatility in production. Accordingly, three functional categories calibrate text difficulty and literacy demands:

Informational Transmission (A1): Texts prioritizing knowledge conveyance and data accessibility;

Literary Aesthetics (A2): Texts foregrounding rhetorical figuration to cultivate empathic attunement to authorial affect, stance, and values;

Specialized Academic (A3): Texts encoding extra-curricular, discipline-specific knowledge systems.

Informational texts, embedding life-world knowledge, present lower processing demands. Literary-aesthetic texts activate rhetorical analysis and critical position-taking, engaging Bloom's higher-order cognition—though L1 literary schemas may partially scaffold interpretation. Specialized-academic texts constitute the apex of difficulty: technical lexicon and epistemic frameworks beyond curricular boundaries invalidate prior knowledge, demanding pure inferential construction. Functional classification necessitates orchestrated analysis of structural patterning, linguistic markedness, content schemata, reader positioning, authorial purpose, and sociocultural embedding. Multifunctionality is the norm; hybridization compounds complexity. The 2024 National Gaokao Reading Section B exemplifies this: ostensibly informational in preferring print dictionaries over digital, it simultaneously deploys self-deprecating “old-fashioned” to aestheticize tactile experience, and contrastive lexis (“pleasure/rewarding” vs. “dull”) to valorize lexicographic labor, thus fusing informational, aesthetic, and specialized (publishing studies) functions.

Accurate functional diagnosis demands: Intercultural literacy to model text-based expectations; Sociological knowledge of institutional structures and issue dynamics; Rhetorical competence to decode figuration and infer latent functions; Deep linguistic processing of lexico-grammatical patterning; Contextual analysis of discursive purpose and audience design; Critical thinking to evaluate logical architecture and ideological positioning; Affective engagement with constructed personas and value orientations.

These operational competencies map precisely onto the academic quality specifications of the *New Curriculum Standards*^[1]. As stipulated, academic quality depicts integrated achievement across core literacy domains: language ability (discourse processing), cultural consciousness (interpersonal positioning), and thinking quality (critical evaluation). The resultant knowledge-practice nexus defines both instructional objectives and assessment constructs, mandating that English pedagogy transcend linguistic formalism to engage contextualized pragmatics, sociocultural application, higher-order cognition, and aesthetic appreciation.

2.1.2 Discursive Scope

Discursive scope—operationalized as text length—functions as a critical mediator of cognitive load and processing depth. While extended texts demand greater temporal investment and activate broader knowledge networks, it is precisely such thematically-sustained, authentic discourses that catalyze the comprehensive reflection necessary for sophisticated meaning-making^[2]. In high-stakes assessment contexts, length intensifies retrieval complexity: expanded information matrices and diffusely distributed relevant data points impede rapid answer localization. Consequently, this sub-dimension calibrates difficulty against the curriculum's three-tiered proficiency architecture, wherein Level 1

(compulsory courses), Level 2 (selective compulsory courses, serving as the Gaokao benchmark), and Level 3 (elective courses) provide pedagogically-meaningful word-count parameters for material selection and task design.

Length-based difficulty operates through three mechanisms: (1) architectural complexity that obscures macro-structure extraction; (2) informational density that attenuates reliance on linguistic intuition; and (3) amplified time-pressure that may exacerbate test-taker anxiety. Mitigating these effects necessitates cultivating advanced discourse sensitivity—the capacity to rapidly identify structural patterns in informational texts, decode lexico-grammatical marking of affect and stance in literary discourse, and synthesize extra-textual knowledge with specialized academic content. This directly aligns with academic quality mandates for mastery of genre architecture, cohesive devices, information status distribution (given vs. new), and inter-clausal semantic-logical relations, thereby manifesting core literacy in both structural parsing and functional interpretation.

The pedagogical imperative is clear: learners must develop automatized linguistic intuition through expansive exposure to varied text lengths and structural configurations within each functional type. This demands an interdisciplinary, multi-genre thematic cluster approach that systematically rotates discursive scope as a deliberate curriculum variable. The scale therefore operationalizes length across three components: total passage word count, aggregate stem word count, and overall test word count. While the first has received extensive scholarly attention, the latter two—directly impacting physical test density and temporal allocation strategies—remain under-theorized.

Total Passage Word Count (B1): (Note: The scale uses B1, B2, B3, but the original only mentioned B2 and B3. I'll assign B1 to the main passage length as it's the most fundamental.) This component is categorized according to curriculum proficiency levels, establishing clear benchmarks for instructional material difficulty.

Total Stem Word Count (B2): Though item stem design rarely treats length as an explicit difficulty parameter, stem verbosity critically constrains stem comprehension, key information extraction, and distractor elimination in foreign language assessment. Aggregated stem length constitutes a substantial proportion of total test volume, directly influencing strategic time management. The “5-in-7” cloze-elision task and integrated “Language Use” sections exemplify this challenge, wherein stems merge passage and options into unified, densely-worded prompts. This format yields stems significantly longer than discrete-item sections, emerging as one of the most cognitively demanding task types—a reality starkly mismatched with the 2019 standard textbooks, which provide scant practice for such configurations. Gaokao's strategic differentiation of stem length across sections—concise for listening (time-pressured), moderate for conventional reading, and maximal for integrated tasks—necessitates a three-tiered classification of aggregate stem word count.

Total Test Word Count (B3): While cumulative test length may theoretically affect reading speed, its empirical impact is negligible when stem length is independently accounted for. Given standardized formats and pre-familiar instructions, the incremental processing overhead from test-level word count does not constitute a distinct source of difficulty meriting separate analytic weight in this framework.

2.1.3 Rhetorical Architecture

This dimension indexes the cognitive accessibility of a discourse's organizational blueprint—its macro-structure, ideational progression, and cohesive scaffolding. The primary difficulty parameter is architectural regularity: the extent to which information flows linearly (temporally or spatially) versus the degree of disruption via flashback, interpolation, or thematic rupture. Each fragmentation event escalates inferential load, taxing working memory and impeding the construction of a coherent mental model. Difficulty therefore scales inversely with structural cohesion. Irrespective of architectural pattern, all structures serve rhetorical purpose: to optimize information transmission, encode authorial stance and affective positioning, introduce novel epistemic content, or provoke critical-aesthetic response. Crucially, unlike other dimensions, organizational structure itself constitutes a direct object of assessment—whether linear, semi-linear, or non-linear, each configuration demands distinct parsing strategies aligned with the author's communicative design.

This dimension operationalizes the *New Curriculum Standards*' emphasis on discourse structure mastery, cohesive device recognition, and logical sequencing of experiential or propositional content. While overlapping partially with text type and length, it more precisely targets the ability to infer authorial intent, distill thematic salience, map logical relations between claims and evidentiary support, decode rhetorical strategizing, and evaluate structural felicity—all constituting deep structural competencies within core literacy.

Pedagogically, instruction must systematically develop students' architectural literacy: in receptive tasks, learners must learn to "read" organizational patterning as metatextual information; in productive tasks, they must strategically deploy structural resources to craft compelling, coherent meanings. The scale differentiates three levels of structural cohesion:

Canonical Architecture (C1): Macro-structure is clearly demarcated (orientation–complication–resolution) with explicit topic sentences, well-formed paragraphing, and overt cohesive signaling (transitional markers, parallel structures), yielding a linear, low-inference processing path.

Latent Architecture (C2): Macro-structure is partially obscured, requiring readers to infer paragraph-level main ideas from supportive detail. Sentence-level relations are implicit, demanding reconstructive inference to establish coherence.

Fragmented Architecture (C3): Non-linear organizational patterns dominate (e.g., anachronic narration, embedded episodes, thematic discontinuity). Readers must perform top-down reassembly, tracking multiple temporal lines or parallel thematic strands to synthesize holistic meaning.

2.2 Background Knowledge

Background knowledge functions as a critical moderating variable in literacy performance, comprising the epistemological schemas—both specialized and general—activated during discourse processing. These schemas cluster into three domains differentiated by their proximity to lived experience and accessibility for transfer: **situational knowledge** (procedural fluency in everyday life-worlds), **disciplinary knowledge** (systemic understanding of scientific and academic domains), and **cultural knowledge** (value-laden frameworks for aesthetic and ethical meaning-making). A test-taker's activated schema repertoire directly determines predictive inference capacity, thematic resolution accuracy, and depth of stance interpretation. While individual variation in experiential exposure is axiomatic, civilizational commonalities render situational knowledge quasi-universal through recurrent daily practice, establishing a baseline processing advantage. Disciplinary knowledge, by contrast, erects vertical barriers between knowledge communities through specialized epistemologies and terminologies. Cultural knowledge occupies an intermediate position: though rooted in life-worlds, it requires formal mediation and symbolic interpretation, rendering it less universal than situational knowledge yet more horizontally transferable than disciplinary expertise, particularly via synergies with humanities curricula (Chinese language arts, history, geography) and acculturated intuition. This epistemological stratification generates clearly demarcated difficulty gradients within Gaokao background knowledge demands.

These knowledge domains operationalize the academic quality standards with distinctive functions. Situational knowledge scaffolds comprehension of "everyday discourse" and "familiar topics" while enabling multimodal text interpretation. Cultural knowledge underpins comparative analysis of Chinese and global cultural heritage, decoding of identity positionings, and interpretation of attitudinal-affective markers—directly manifesting intercultural competence. Disciplinary knowledge facilitates processing of cross-disciplinary conceptual vocabulary and specialized terminology. Collectively, these domains activate two core competency clusters: (1) linguistic capacity for multimodal meaning inference, and (2) cultural consciousness for perceiving, analyzing, and ethically evaluating cultural variation while maintaining global-mindedness. These competencies constitute the substrate for enhanced learning capability.

The richness of background knowledge positions English learning not as an insular subject discipline but as a dynamic interface with contemporary society's transcultural spatiality^[3] and interdisciplinary knowledge production. Pedagogically, this mandates that English education transcend linguistic formalism to embrace the active co-construction of knowledge within expanded epistemic horizons.

Situational Knowledge (D1): Encompasses the procedural schemas and practical competencies requisite for life-world navigation, including socio-pragmatic protocols (ceremonial etiquette, interactional turn-taking), self-regulatory practices (hygiene maintenance, health surveillance), civic literacy (fundamental legal rights consciousness, civil participation frameworks), domestic operational skills (culinary procedures, spatial organization, financial budgeting), and risk mitigation competencies (first aid protocols, fire suppression and security measures).

Cultural Knowledge (D2): Constitutes the symbolic systems, heritage artifacts, and value-orientational frameworks encoding collective meaning within particular cultural formations,

spanning language varieties, literary canons, historiographic narratives, aesthetic traditions, cosmological paradigms, and ritualized customary practices.

Disciplinary Knowledge (D3): Comprises specialized epistemological frameworks—foundational constructs, axiomatic principles, methodological protocols, and applied competencies—within circumscribed knowledge domains. Notably, in basic education, disciplinary knowledge rarely manifests in pure form; rather, it emerges entangled with situational and cultural knowledge through anchoring in everyday phenomena, socio-historical matrices, and cultural artifacts, generating hybridized knowledge complexes that resist categorical purification.

2.3 Linguistic Knowledge Architecture

This dimension operationalizes the depth and integration of lexical-grammatical resources requisite for meaning-making in problem-solving contexts. For Chinese learners, the systemic divergence between English and Chinese morphosyntactic and lexical systems constitutes a primary processing constraint. Proficiency differentiation in high-stakes assessment hinges on the degree to which candidates must mobilize multi-layered knowledge to infer discourse meaning, particularly the interplay between item stems and co-textual/contextual cues that determines efficient answer location and construct representation. Consequently, the embeddedness and convergence of lexico-grammatical patterning, phraseological units, and syntactic configurations within stems and their radiating discourse directly mediate item difficulty and index competency level. This architectonic perspective—where lexical, phrasal, clausal, and textual structures form an interdependent, mutually-reinforcing system—implies that test-based meaning is irreducibly compositional. No foreign language assessment item can be monolithic; each activates at least two knowledge domains, with lexical and grammatical resources acquiring context-specific semantic prosodies through the test-writer's rhetorical design. This accords with academic quality specifications mandating context-sensitive grammatical selection, vocabulary connotation/denotation resolution, implicature recovery, appreciation of marked linguistic choices, and culturally-appropriate expression—all of which must be re-enacted in productive tasks through strategic lexico-grammatical orchestration aligned with intended perlocutionary effects.

The competency correspondence is fundamentally integrative: deploying linguistic resources to transmit information, encode personal stance and affect, signal intentionality and values, and construct appropriate identities and interpersonal roles. Embedded within this dimension is the cognitive capacity to leverage linguistic knowledge for epistemic advancement—inferring logical and associative relations among propositions, inducing conceptual categories, and deploying deductive/explanatory reasoning to novel problem spaces.

English pedagogy, like all language education, is communication-principled: rhetorical purpose and expressive effect drive lexical selection and textual organization. The contextually-nuanced meanings of vocabulary, syntax, and grammar—plus their ideational and interpersonal effects—constitute the crux of foreign language praxis. Instruction is thus a meaning-driven process anchored in linguistic mastery, oriented toward intercultural comparison, and aimed at communicative efficacy.

2.3.1 Knowledge Density

This sub-dimension quantifies the number of distinct lexical-grammatical resources a candidate must orchestrate to process the stem and select the correct response. Items may isolate single features (e.g., part-of-speech conversion, subordinator selection, semantic field differentiation, tense marking) or demand concurrent activation of multiple systems. Required resource integration intensifies commensurately with difficulty.

Single-Point Activation (E1): Resolution draws on one discrete linguistic feature.

Dual-Point Integration (E2): Resolution requires coordinated deployment of two linguistic systems.

Multi-Point Synthesis (E3): Resolution necessitates simultaneous orchestration of multiple linguistic resources.

2.3.2 Semantic Strata

Grammatical and lexical meaning potentials stratify into three levels: formal, contextual, and rhetorical. Formal meaning denotes literal, decontextualized sense. Contextual meaning captures situated extension and connotational enrichment triggered by co-text. Rhetorical meaning emerges when contextual extensions acquire figurative or interpersonal functions (metaphor, metonymy, irony),

accruing aesthetic-literary value. From formal through contextual to rhetorical strata, the discourse span of required knowledge widens, processing depth increases, and difficulty escalates correspondingly.

Formal (F1): Resolution targets literal form-meaning mapping.

Contextual (F2): Resolution demands inference of situated, co-textually-modulated meaning.

Rhetorical (F3): Resolution requires interpretation of figuratively-charged, aesthetically-marked meaning.

2.3.3 Lexical Breadth

Vocabulary size fundamentally constrains both receptive precision and expressive sophistication. Since Level 2 of the academic quality standards anchors Gaokao difficulty, and textbook sequences map directly onto proficiency tiers, vocabulary thresholds align accordingly. Regional additions of ≤ 200 items are subsumed within the Elective tier for metric consistency.

Core Vocabulary (G1): Lexis from Required Courses ($\leq 1,500$ tokens)

Extended Vocabulary (G2): Lexis from Selective Required Courses ($\sim 2,000$ tokens)

Specialized Vocabulary (G3): Lexis from Elective Courses ($\sim 3,000$ tokens)

2.4 Knowledge Mobilization Pathways

This dimension operationalizes the cognitive trajectories through which test-takers access, reconstruct, and evaluate evidentiary bases to generate warranted responses. As a discourse-based proficiency test, the Gaokao calibrates difficulty through the dispersal of evidentiary sources: the wider the distribution, the greater the demand for knowledge orchestration beyond mere processing volume. This necessitates not only linguistic decoding but also the strategic activation of prior knowledge, comparative analysis, imaginative hypothesizing, logical inference, and evaluative synthesis. When evidentiary sources transcend textual boundaries, migrating into test-takers' experiential repertoires and aesthetic sensibilities, the cognitive register shifts toward innovative meaning-making and deep critical reflection, escalating difficulty correspondingly. The dimension stratifies epistemic access into three hierarchical levels:

Direct Retrieval (H1): Explicit answers locatable within a single, contiguous textual source requiring minimal inferential bridging.

Synthesized Inference (H2): Answers constructed through integrative reasoning across multiple, discrete information loci within the discourse.

Evaluative Reconstruction (H3): Answers requiring the full orchestration of extra-textual prior knowledge coupled with aesthetic appreciation, critical interrogation, and creative knowledge application.

These mobilization pathways implicate a constellation of competencies—discourse parsing, cross-cultural positioning, higher-order cognition, and metacognitive regulation—directly mirroring the academic quality standards' specifications: comprehending literal and implied meaning, articulating and justifying positions, executing comparative-analytical-synthetic operations, and performing evaluative-appreciative judgments. In essence, the dimension captures the full scope of epistemic work, from critical examination of propositional content and ideological stance to appreciation of linguistic artistry, from raising warranted skepticism to constructing autonomous value judgments.

The assessment of learning capacity—particularly its metacognitive stratum—has long posed measurement challenges. However, the *China College Entrance Examination Evaluation System* operationalizes “learning mastery” through three measurable sub-constructs—information acquisition, comprehension mastery, and knowledge integration—thereby translating abstract metacognitive processes into observable performance indicators^[4]. This concretization affirms that foreign language education is fundamentally a process of integrated competency cultivation, wherein academic quality descriptors become direct specifications for targeted pedagogical design and assessment architecture.

2.5 Text Readability

This dimension operationalizes an adapted Flesch Reading Ease metric to quantify micro-level processing demands imposed by lexical-syntactic surface features^[5]. While higher-order

comprehension—integrating background knowledge, inferential reasoning, and evaluative judgment—constitutes the primary construct of literacy assessment, surface complexity exerts measurable influence on processing efficiency and working memory allocation. The metric functions as a convergent validity check, triangulating algorithmic indices with qualitative difficulty judgments derived from preceding dimensions.

Flesch Reading Ease Formula:

Flesch Reading Ease

$$= 206.835 - 1.015 \times (\text{Average Sentence Length}) - 84.6 \times (\text{Average Syllables per Word}) + 1.5 \times (\text{Syntactic Subordination Index})$$

Average Sentence Length: Total words $\div$ total sentences

Average Syllables per Word: Total syllables $\div$ total words

Syntactic Subordination Index: Subordinate clauses $\div$ total clauses (0-1 scale, measuring clausal embedding complexity)

Note: The final term adapts the classical Flesch formula by incorporating syntactic depth as a complexity multiplier, reflecting research demonstrating that subordination density significantly attenuates reading fluency beyond word- and sentence-level metrics.

Flesch-Kincaid Grade-Level :

Low Processing Load (I1): Score 70-100 (accessible to Level 1 proficiency)

Moderate Processing Load (I2): Score 60-70 (target Level 2 proficiency)

High Processing Load (I3): Score 0-60 (demanding Level 3 proficiency)

These thresholds align with the curriculum standards' explicit division of academic proficiency into three developmental stages, enabling direct mapping between surface feature quantification and expected learner processing capacities.

3. Implications of Scale Design for High School English Teaching, Learning, and Assessment

In contemporary English education, characterized by a shift from testing culture to assessment culture, all evaluations are learner-centered and learning-centered. Designing a discipline literacy gradient scale based on the New College Entrance Examination precisely visualizes assessment and test design, thereby facilitating daily teaching and assessment that promote learning^[6].

From the Perspective of Discourse Environment: High school English teaching should implement thematic cluster text teaching based on themes. Consequently, assessment should focus on learners' knowledge of and ability to analyze text features, stylistic characteristics, discourse connotation, text organization, and text comparison across different types of texts within thematic clusters.

From the Perspective of Knowledge Scope: High school English teaching should be interdisciplinary. Assessment should center on the breadth of students' knowledge and dimensional abilities in applying prior and new knowledge to grasp themes, analyze main ideas and details, reason, evaluate and interpret texts, and assess social phenomena.

From the Perspective of Language Skills Application: High school English teaching is an integrated approach where the skills of viewing, listening, speaking, reading, writing, and translation intersect and mutually enhance one another. Assessment should examine students' ability to acquire relevant knowledge and information through multimodal resources, organize, analyze, understand, and evaluate the producer's viewpoints, emotions, attitudes, and values, and express their own viewpoints, emotions, attitudes, and values.

From the Perspective of Disciplinary Connotation: High school English teaching should be an integrated approach to disciplinary literacy that merges knowledge, skills, and thinking, characterized by holistic complexity. Consequently, assessment is a test of test-takers' abilities to understand discourse, critically interpret discourse, and use prior knowledge and discourse knowledge to communicate with the outside world and understand the world.

From the Perspective of Disciplinary Educational Value: High school English teaching is

cross-cultural instruction. Assessment should focus on abilities such as the application of communication strategies, cultural understanding, cultural criticism, cultural rationality, and cultural learning.

4. Conclusion

The Literacy Gradient Analysis Scale constructed in this paper marks a significant breakthrough in transitioning English basic education assessment from fragmentation to systematization. Its multidimensional framework—encompassing discourse form, background knowledge, language knowledge, answer retrieval complexity, and text readability—not only quantifies the literacy differentiation mechanisms of the New College Entrance Examination's authentic test items, but also deeply informs the core transformative direction of high school English teaching. Empirical evidence from the scale reveals that the hierarchical design of test questions in discourse types (e.g., from information transmission A1 to specialized learning A3) and answer retrieval pathways (H1 to H3) directly echoes the academic quality standards' requirements for higher-order thinking (such as critical interpretation and cross-cultural understanding), underscoring the urgency of shifting teaching from knowledge transmission to competency integration. For instance, the analytical indicators of discourse length (B2-B3) and organizational structure (C1-C3) demonstrate that proficiency in processing lengthy texts depends on the cultivation of language intuition and interdisciplinary clustered text teaching. This necessitates that teaching materials and classroom activities transcend traditional genre limitations, strengthening training in structure identification and information location.

Based on the scale's implications, the high school English teaching, learning, and assessment system requires deepened reform across five dimensions: First, the discourse environment dimension demands that teaching conduct comparative multi-text analyses around themes, with assessment focusing on genre features and connotative analysis to enhance students' discourse discernment and logical reasoning abilities. Second, the knowledge scope dimension advocates for interdisciplinary integration, with assessment needing to examine students' ability to interpret social phenomena using both prior and new knowledge. Third, the language skills application dimension emphasizes the integration of viewing, listening, speaking, reading, writing, and translation, with assessment relying on multimodal resources (such as videos and AI tools) to design contextualized tasks that enhance students' information processing and affective expression effectiveness. Fourth, the disciplinary connotation dimension requires teaching to integrate knowledge, skills, and thinking, with assessment focusing on the depth of discourse comprehension. Fifth, the disciplinary educational value dimension points toward cross-cultural communication, with assessment incorporating cultural understanding and strategy application to cultivate global competence.

Future research could further explore the synergistic innovation of the scale with technology, for example, by integrating AI-driven personalized learning platforms to optimize training in answer complexity, or by drawing on data literacy models to enhance the real-time capability and scalability of assessment. In conclusion, this scale provides English education with a scientific assessment tool and pedagogical blueprint. Through continuous iteration and practical validation, it will propel basic education in the core competencies era toward greater equity, quality, and sustainability.

Acknowledgments

Research and Practice on Senior High School English Teacher Education Based on Academic Quality Monitoring (Yue Jiao Gao Han [2023] No. 29), Research and Practice Project on New Normal Education Construction Promoting High-Quality Development of Basic Education;

Research on the Adaptability of English Teachers in Western Guangdong to the New Curriculum Standards for Compulsory Education (2023GXJK382), Guangdong Provincial Education Science Planning Project;

Research on the Thresholds in Alice Munro's Works (ZW2021016), Humanities and Social Sciences Talent Program of Lingnan Normal University.

References

[1] Ministry of Education of the People's Republic of China. *Senior high school English curriculum*

- standards (2017 edition, 2020 revision)[M]. Beijing: People's Education Press, 2020: 46.
- [2] Cheng X T. Core Competencies of English Discipline and Their Assessment[J]. *China Examinations*, 2017, (5): 7–14.
- [3] Liu Y M. *Big Culture Theme-Based Teaching: Rethinking Middle School English Teaching*[M]. Suzhou: Suzhou University Press, 2024: 75.
- [4] Ministry of Education Examination Center of the People's Republic of China. *China's Gaokao Evaluation System*[M]. Beijing: People's Education Press, 2019: 19.
- [5] Imperial J M, Madabushi H T. Flesch or Fumble? Evaluating Readability Standard Alignment of Instruction-Tuned Language Models[EB/OL]. [2024-11-21]. <https://sc.panda985.com/scholar?q=Flesch%20or%20Fumble%3F%20Evaluating%20Readability%20Standard%20Alignment%20of%20Instruction-Tuned%20Language%20Models>.
- [6] Lin D L. *A Reference Framework for Language Assessment Literacy of Primary and Secondary School English Teachers*[M]. Beijing: Foreign Language Teaching and Research Press, 2018.