Фундаментальные основы и междисциплинарные связи / Fundamental Principles and Interdisciplinary Connections

Научная статья
Original article
DOI: 10.12731/2658-4034-2026-17-2-1124
EDN: XUYBHK
UDC 372.881.111.1

Analysis of the alignment between English graduation exam tasks and educational standards in the context of school education reform in China

Wu Xi, Xie Zhongyuan
College of Foreign Languages, Xinjiang Normal University, Urumqi, China

Abstract

Background. The article examines misalignment between English graduation exam tasks and national curriculum standards during China's competency-based education reform. Curricular focus on communicative competence and higher-order thinking contrasts with exam tasks that emphasize surface knowledge, particularly in listening and reading, potentially hindering meaningful language learning.

Purpose. This study evaluates how well English exam tasks match the 2022 Compulsory Education English Curriculum Standards, using the Webb Alignment Model to identify content and cognitive gaps.

Materials and methods. The study is based on an analysis of the 2024 English graduation examination administered in the Xinjiang Uyghur Autonomous Region. A four-dimensional framework, including knowledge category, depth of knowledge, breadth of coverage, and balance of distribution, was used to measure alignment. Cognitive levels were classified using an integrated Bloom Webb taxonomy. The study also included semi-structured interviews with English teachers to supplement quantitative findings with contextual insights.

Results. The analysis revealed acceptable alignment in the knowledge category and cognitive complexity, with 89.47% of tasks in the Language Knowledge domain meeting or exceeding expected cognitive levels. However, the exam demonstrated poor alignment in terms of breadth of coverage and distributional balance, particularly in the Language Skills domain, where only 28.27% of objectives were addressed, and the balance coefficient was below the acceptable threshold. Teacher interviews highlighted the influence

of practical constraints, such as scoring feasibility and lack of training, on task design. These findings show that even though current assessments partially reflect curriculum standards, there are still significant gaps in assessing communicative and higher-order competencies. The study also concluded that if one wants to improve the consistency of exams, it is necessary to enrich the question types, broaden the coverage of course objectives, and enhance teachers' assessment skills. These steps are crucial to ensure the final assessment accurately reflects the goals of competency-based education and supports the development of students' communication skills.

Keywords*:* curriculum alignment; English exam; Webb Alignment Model; cognitive complexity; communicative competence; China education reform

For citation. Wu, X., & Xie, Z. (2026). Analysis of the alignment between English graduation exam tasks and educational standards in the context of school education reform in China. *Russian Journal of Education and Psychology*, *17*(2), 100–121. https://doi.org/10.12731/2658-4034-2026-17-2-1124

Анализ соответствия заданий выпускного экзамена по английскому языку образовательным стандартам в условиях реформы школьного образования в Китае

У Си, Се Чжунъюань
Институт иностранных языков, Синьцзянский педагогический университет, Урумчи, Китай

Аннотация

Обоснование. В статье рассматривается проблема несоответствия между заданиями итогового экзамена по английскому языку и требованиями национального учебного стандарта в контексте перехода системы образования Китая к компетентностной модели. Несмотря на то, что новая программа подчеркивает важность формирования коммуникативной компетенции и развития навыков высокого уровня мышления, содержание многих экзаменационных заданий по-прежнему сосредоточено на поверхностных знаниях и буквальном понимании, особенно в разделах аудирования и чтения. Подобное несоответствие может снижать эффективность оценки и препятствовать достижению реальных образовательных целей.

Цель. Цель исследования определить степень соответствия заданий экзамена по английскому языку требованиям стандарта «Обязательного образования по английскому языку» (редакция 2022 г.) с применением модели согласования Уэбба и выявить содержательные и когнитивные расхождения.

Материалы и методы. В исследовании проведён анализ итогового экзамена по английскому языку 2024 г. для учащихся средней школы Синьцзян-Уйгурского

автономного района. В качестве инструмента анализа использована четырехмерная модель Н.Л. Уэбба: категория знаний, глубина знаний, широта охвата и сбалансированность распределения заданий. Классификация уровней когнитивной сложности основана на интеграции таксономии Блума и модели Уэбба. Для дополнения количественных данных были проведены полуструктурированные интервью с учителями английского языка.

Результаты. Анализ показал приемлемый уровень согласования по категориям знаний и когнитивной сложности: 89,47% заданий в области языковых знаний соответствовали или превышали ожидаемый уровень. Однако по таким показателям, как широта охвата и сбалансированность распределения, особенно в области языковых навыков, было зафиксировано значительное несоответствие: охвачено лишь 28,27% целей, а коэффициент распределения оказался ниже допустимого порога. Интервью с преподавателями выявили влияние практических ограничений таких как трудности объективной оценки и нехватка методической подготовки на проектирование заданий. Таким образом, несмотря на частичное соответствие экзамена стандарту, остаются серьезные пробелы в оценке коммуникативной и когнитивно сложной деятельности обучающихся. В заключение подчеркивается необходимость расширения типов заданий, увеличения охвата целей учебной программы и повышения оценочной грамотности педагогов для более точного отражения задач компетентностного образования и поддержки развития коммуникативной компетентности школьников.

Ключевые слова: согласование стандартом; экзамен по английскому языку; модель Уэбба; когнитивная сложность; коммуникативная компетенция; реформа образования Китая

Для цитирования. У, Си, & Се, Чжунъюань (2026). Анализ соответствия заданий выпускного экзамена по английскому языку образовательным стандартам в условиях реформы школьного образования в Китае. *Russian Journal of Education and Psychology*, *17*(2), 100–121. https://doi.org/10.12731/2658-4034-2026-17-2-1124

Introduction

In the context of ongoing global transformations and China's nationwide shift toward competency-based education, the role of scientifically grounded student assessment has become increasingly critical [11]. Central to these reforms is the alignment between curriculum goals and the structure of final examinations, which serve not only as tools of measurement but also as mechanisms guiding instructional focus and learning outcomes [11]. The alignment challenge is especially serious in foreign language education. This field requires assessment that goes beyond testing students' memo-

ry of facts. The goal of assessment is to measure students' communicative competence. Content subjects usually focus on factual knowledge. In contrast, language subjects need to assess how well students can use language in real situations. This kind of evaluation involves complex and integrative tasks. These tasks require students to show their ability to use language for communication in different contexts. Test designers face both theoretical and practical challenges. Based on that it is necessary to create exams that reflect these various language abilities. These difficulties make English language assessment very important in China. It serves as a key example of how China's educational reforms are working in practice.

The relevance of this study lies in the need for a comprehensive approach to design and evaluation of examination materials that ensures coherence between goals, content, and methods of instruction. The proposed research focuses on identifying the degree of content and cognitive alignment between English graduation exam tasks and the established educational standards, which holds both theoretical and practical significance for experts in pedagogy, foreign language teaching methodology, and educational assessment.

The *purpose* of this article is to conduct a comprehensive analysis of the degree to which English graduation exam tasks align with the requirements of the national educational standard of the People's Republic of China, within the context of the ongoing general secondary education reform. The study aims to identify the content, cognitive complexity, and orientation of exam tasks in comparison with the goals and learning outcomes outlined in the current instructional and regulatory documents, in order to evaluate their pedagogical validity and their consistency with modern approaches to the development and assessment of students' foreign language communicative competence.

Literature review on the problem

In recent years, the alignment between educational standards and final examinations has become a central focus of pedagogical research, especially in the context of competency-based educational reform. A key goal is to examine how well assessment tasks reflect the learning outcomes prescribed in national curricula, particularly in foreign language education.

A strong theoretical and methodological foundation for such studies is provided by international research on standard transformation in the era of globalization and digitalization. One of the most influential frameworks is Webb's Depth of Knowledge (DOK) model, which offers a multidimensional approach to evaluating the cognitive complexity and alignment of exam tasks with learning standards [13]. M.W. Kirst further contributes to this discussion by outlining practical methods for aligning curriculum content with assessment tasks in standards-based systems [3]. These models are broadly compatible with the principles of competence-oriented curriculum design and form the methodological basis for the current study. At the conceptual level, scholars like A.V. Khutorskoy and A.A. Malygin emphasize that competency-based education demands a reassessment of traditional grading systems and evaluation formats in favor of approaches that prioritize meta-subject and transferable skills [2; 7].

In the Chinese and Russian academic contexts, research has also focused on identifying gaps between curriculum goals and exam design. Yan Yinlu and colleagues highlight the urgency of aligning exam structures with reformed secondary education standards in China [15]. Liu Jie directly points to the mismatch between what is taught and what is tested, noting the negative impact on validity and fairness [5]. Tang Wenwen's comparative analysis of language testing practices in Russia and China identifies key methodological differences, particularly in how test items are developed in relation to curricular goals [14]. Similarly, G.S. Kovaleva and K.A. Krasnyanskaya examine best international practices, concluding that alignment between assessment and standards is a major determinant of overall educational quality [4].

At the regional and implementation level, more practice-oriented studies offer insight into the operational challenges of maintaining alignment. For instance, N.V. Maizenger's evaluation of English exam results in Altai Krai shows how task format and design affect both student outcomes and compliance with federal standards [6]. N.N. Naidenova emphasizes the need for systemic review of assessment complexity, especially when curriculum expectations rise faster than test sophistication [9]. T. B. Volobueva underscores the role of teacher training in this process, suggesting that digital integration models can help synchronize assessment practices with

new curricular frameworks [10]. Finally, broader historical analyses such as the work of S.E. Mansurova and K.A. Tabarovskaya provide a useful background by tracing the evolution of curriculum standardization in educational policy [8].

In summary, the reviewed literature highlights three main issues: the theoretical need for alignment between cognitive depth and assessment structure; the practical difficulties of achieving this alignment within high-stakes exam systems; and the importance of teacher preparedness and assessment literacy in supporting the goals of curriculum reform. These themes collectively inform the methodological approach and empirical focus of the present study, which aims to evaluate the degree of alignment between English exam tasks and curriculum objectives in China.

Problem statement. Amid China's ongoing education reform, the effectiveness of final assessments depends on their ability to reflect the competencies outlined in national standards. However, current English examination tasks often fall short in evaluating key communicative skills such as pragmatic awareness and inferential reasoning. For example, listening sections rarely assess a student's ability to interpret speaker intent, while reading tasks tend to focus on factual recall rather than discourse-level understanding. This misalignment between exam content and curricular goals undermines the validity of the assessment and may distort classroom instruction. The core issue lies in the lack of a systematic evaluation of how well exam tasks represent the cognitive and content demands of the curriculum.

The *scientific novelty* of this study lies in the application of the Webb Alignment Model to systematically evaluate the alignment between English graduation exam tasks and the 2022 national curriculum standards in China, a method not yet widely applied in the Chinese context. By combining cognitive complexity analysis with task content coding, the study reveals hidden misalignments between exam practices and curriculum goals, particularly in the assessment of higher-order communicative competencies.

The *theoretical significance* of the article lies in its contribution to the modeling of curriculum-assessment alignment in the field of foreign language education. It provides a structured framework for analyzing how assessment content and cognitive demands reflect competence-based educational

objectives. The findings offer a reference point for comparative studies of certification systems and support ongoing efforts in aligning pedagogical design with national educational reform strategies.

The *practical significance* of the research is determined by the potential use of its findings in improving the practice of developing graduation exam tasks, as well as in the professional training of teachers and methodologists in the English language instruction. The conclusions presented can support the creation of new assessment tools aligned with the competence-based models adopted in China's contemporary education strategies.

Thus, this study helps establish the degree of alignment between exam content and educational standards, an essential step toward enhancing the objectivity of final assessment and ensuring the quality and effectiveness of language education under systemic reform.

Materials and methods

The object of this study is the set of English language examination tasks included in the 2024 final assessment of junior secondary school students in the Xinjiang Uygur Autonomous Region of the People's Republic of China, as well as the provisions of the Compulsory Education English Curriculum Standards (2022 Edition). The research aims to identify the degree of alignment between the content and cognitive complexity of the examination tasks and the educational goals defined in the updated national standard, which is grounded in a competence-based approach.

The methodological foundation of the research is based on the Webb Alignment Model (N.L. Webb), adapted to the context of the ongoing educational reform in China. According to this model, the correspondence between educational standards and assessment materials is measured through four dimensions: categorical relevance of knowledge, depth of knowledge, breadth of coverage, and balance of distribution. Although originally developed within the U.S. educational context, the Webb Alignment Model is methodologically compatible with the logic of China's 2022 Compulsory Education English Curriculum Standards. Both frameworks emphasize a hierarchical progression from knowledge acquisition to comprehension and practical application core principles of

the current competence-based reform. This structural similarity ensures that the model can effectively capture the content and cognitive alignment between exam tasks and curriculum objectives. Each of these dimensions includes a criterion for acceptable and unacceptable levels of alignment (see Table 1).

Table 1.

Analytical Framework of the Webb Alignment Model

Dimension of Alignment	Content of Analysis	Acceptable Level	Unacceptable Level
Knowledge Category	Correspondence between task content and curriculum content	≥ 6 tasks covering the domain	< 6 tasks
Depth of Knowledge	Cognitive complexity level of tasks relative to standards	≥ 50% of tasks meet the standard level	< 50% of tasks
Breadth of Coverage	Extent of learning objectives covered within a domain	≥ 50% of objectives are covered	< 50% of objectives
Balance of Distribution	Evenness of task distribution across objectives	Coefficient ≥ 0.7	Coefficient < 0.7

Source: Produced by the author

To determine cognitive complexity, the study employed a classification combining Bloom's Taxonomy with Webb's depth-of-knowledge levels. All objectives of the curriculum were coded according to three hierarchical levels:

A (Memory) – recall and recognition of facts and rules;

B (Understanding) – interpretation and explanation of content;

C (Application) – use of acquired knowledge in new contexts.

Table 2.

Classification of Cognitive Verbs in the Curriculum Standards

Level	Code	Description	Example Verbs
A (Memory)	A	Recall of factual, procedural, or conceptual knowledge	identify, describe, define, recognize
B (Understanding)	B	Ability to interpret or infer meaning	comprehend, distinguish, infer, explain
C (Application)	C	Application of knowledge to new situations or problem-solving	apply, integrate, evaluate, modify

Source: Produced by the author

The process of coding included all target competencies of the curriculum, from language knowledge (phonetics, vocabulary, grammar, text) to language skills (listening, reading, writing). Each task of the 2024 English examination was matched with the specific curriculum goal and assigned a cognitive code (A/B/C). For instance, the task requiring students to choose the correct pronoun in context was coded as B-1.2.2, while a task demanding logical sentence completion was coded as C-1.4.4.

An important stage of the study was a content analysis of the examination paper, evaluating the correspondence of each task to the curriculum framework. The balance of task distribution was determined using the following formula:

$$I=\frac{\sum(I_k/H)^2}{O}$$

where I is the balance coefficient, I is the number of tasks corresponding to a specific goal, H is the total number of tasks in the domain, and O is the number of curriculum goals covered by the exam. According to Webb's model, a coefficient value of $I \geq 0.7$ is interpreted as an acceptable level of alignment [12].

To conduct the analysis, the study used a three-level coding system based on the structure of "learning domain – thematic goal – specific objective." According to the 2022 Compulsory Education English Curriculum Standards, there are two main learning domains: language knowledge (including phonetics, vocabulary, grammar, and text) and language skills (including listening, reading, and writing). Within these domains, a total of 65 specific learning objectives were identified and grouped based on their level of cognitive complexity. A quantitative summary of all coded curriculum objectives is presented in Table 3. These data provided the reference framework for matching the curriculum codes with the task codes, allowing the measurement of alignment across all four dimensions.

Table 3.

Coding Results Based on the Compulsory Education English Curriculum Standards (2022 Edition)

Learning Domain	Thematic Goals	Specific Objectives	Cognitive Depth		
			Recall – A	Understand – B	Apply – C
Language Knowledge	4	19	7	4	8
Language Skills	3	46	7	19	20

Source: Produced by the author

In addition to quantitative coding, the study incorporated semi-structured interviews with three English teachers and one curriculum expert from Urumqi. These interviews provided insights into practical challenges of exam design and curriculum implementation, and helped contextualize the quantitative findings.

Results and discussion

1. Overall alignment of exam tasks with curriculum goals

The first dimension of analysis is the alignment between exam tasks and curriculum goals. This alignment is evaluated at the level of content categories. The Webb Alignment Model calls this level the Knowledge Category. This dimension evaluates how well the exam tasks reflect the curriculum objectives. These objectives are defined by the themes and topics that the curriculum covers. Webb provides specific criteria for acceptable alignment. According to his model, alignment is acceptable when exam tasks address more than six specific objectives in one learning domain. The results of the analysis are summarized in Table 4.

Table 4.

Knowledge category alignment summary

Learning Domain	Total Curriculum Objectives	Objectives Addressed in Exam	Percentage Covered	Alignment Achieved
Language Knowledge	19	8	42.11%	Yes
Language Skills	46	13	28.27%	Yes

Note: Minimum acceptable threshold = 6 objectives addressed per domain.

As shown in Table 4, the Language Knowledge domain meets the minimum threshold set by the alignment model. On the other hand, the Language Skills domain also meets this requirement. These results suggest that there is an overall acceptable level of alignment. The exam content corresponds reasonably well with the structure of the curriculum. However, the relatively low coverage percentages, 42.11% for Language Knowledge and 28.27% for Language Skills, indicate that a substantial portion of the curriculum objectives remains unaddressed.

The Language Skills domain demonstrates broader thematic engagement, which reflects a more integrated and authentic approach to task design. Tasks

in this domain, particularly in reading and writing, typically require students to comprehend, synthesize, and express information in contextually meaningful ways. This supports the curriculum's emphasis on communicative competence and the practical application of language in real-life scenarios.

In contrast, the Language Knowledge domain, while achieving formal alignment, exhibits a more selective focus. The current examination has reduced reliance on isolated vocabulary or grammar questions and shifted toward tasks that embed language knowledge within broader discourse contexts. This change is aligned with recent educational reforms. These reforms are aimed at encouraging deeper language processing instead of surface-level memorization. For example, grammar knowledge and lexical knowledge are now often assessed through integrated reading tasks or cloze activities. These tasks require students to use both recognition skills and application skills.

However, data shows that coverage in both areas is low. This means that the focus of the content has been narrowed. Some subdomains, such as phonetic knowledge or pragmatic competence, are easily overlooked or neglected. This is mainly due to the limitations of written assessments in capturing these competencies. This issue is especially noticeable in the Language Skills domain. In this domain, complex learning goals such as interpreting the speaker's intent or expressing nuanced opinions are often simplified. These goals are frequently reduced to tasks that focus only on retrieving factual information.

From a pedagogical perspective, the narrowing of the scope of exam content may stem from several contextual constraints. Firstly, the curriculum itself is broad in scope, especially within a competency-based framework. This presents challenges for summative assessment. Except that test designers often prioritize task reliability and the operability of scoring. To ensure consistent scoring standards and an efficient scoring process, some open-ended tasks or those involving higher-order thinking may be intentionally excluded. Finally, regional factors may also influence test design choices. For example, some regions have relatively limited assessment experience, and educational resources are unevenly distributed. These problems are particularly pronounced in educationally underdeveloped areas.

In summary, the exam met the minimum standards of category consistency in both learning areas, validating its basic compliance with curriculum requirements. However, the scope of the exam objectives was relatively limited, and the task design requires further improvement. A more balanced distribution across curriculum categories would enhance not only the content validity of the assessment but also its ability to support meaningful language learning outcomes in line with national education reform goals.

2. Cognitive complexity alignment

The second alignment dimension focuses on the consistency between the cognitive complexity of exam tasks and the expected levels defined in the national curriculum standards. Within the framework of the Webb Alignment Model, this dimension evaluates whether assessment items meet the depth-of-knowledge expectations corresponding to the cognitive demands of learning objectives. The threshold for acceptable alignment is met when at least 50% of the tasks correspond to or exceed the curriculum's specified cognitive levels. The results of the alignment analysis are presented in Table 5.

Table 5.

Cognitive Complexity Alignment Summary

Learning Domain	Total Exam Tasks Analyzed	Tasks Below Cognitive Level	Tasks Aligned	Tasks Above Cognitive Level	Alignment Ratio (%)	Alignment Achieved
Language Knowledge	38	4	20	14	89.47%	Yes
Language Skills	58	28	26	4	51.72%	Yes

Note: Alignment threshold ≥ 50% (aligned + above-level tasks).

The results indicate that both learning domains reach the minimum acceptable threshold for cognitive alignment. In the Language Knowledge domain, nearly 90% of tasks meet or exceed the expected level of cognitive demand, reflecting a strong correspondence with the curriculum. This is largely due to the nature of knowledge-focused items, which tend to concentrate on recognition, identification, and structured transformation of form objectives that are explicitly outlined and relatively straightforward to assess within controlled formats.

In contrast, the Language Skills domain exhibits a marginal alignment level (51.72%), with a substantial number of tasks falling below the required cognitive level. These tasks typically involve surface-level processing, such as retrieving factual information from listening or reading passages. For example, many listening items assess literal comprehension without requiring inference, synthesis, or contextual interpretation. Similarly, reading items often ask students to locate explicit details rather than evaluate arguments or analyze textual structure.

The discrepancy between the two domains can be attributed to differences in the cognitive design of learning objectives. In the Language Knowledge domain, the curriculum primarily emphasizes low to mid-level cognitive operations, such as describing grammatical features, identifying word forms, or understanding sentence structure. These expectations are relatively easier to fulfill through conventional testing formats.

However, the curriculum's expectations for Language Skills involve higher-order thinking, including inference, integration of multimodal information, and the ability to evaluate or generate text in context. Despite these elevated objectives, many of the corresponding exam tasks remain limited to lower-level operations. As a result, there is a noticeable gap between the intended depth of knowledge and the actual cognitive demands of the test.

The design of tests is greatly influenced by the educational environments of the region. In areas with limited resources, like Xinjiang, the creation of exams typically prioritizes straightforwardness and basic abilities evaluation. Consequently, these tests tend to be more traditional, emphasizing dependability and equity, often neglecting more complex cognitive tasks.

Even though they appear to align with the curriculum's themes, tasks focusing on language skills often struggle to understand the curriculum's fundamental cognitive goals, raising questions about the efficacy of evaluation techniques in curricular structures centered on competencies.

If students consistently fail to reach the necessary cognitive complexity in testing tasks, their intellectual engagement may decrease, thus hindering the development of key skills such as analysis and problem-solving ability. Moreover, over-reliance on basic tasks may limit class instruction and restrict student opportunities to use the language in real time.

To address this imbalance, it is important to target professional development, which is needed for test design, particularly for listening and reading comprehension tasks. Incorporating types of items that aid in interpretation, synthesis, and evaluative reasoning can bridge the division between the objectives of the curriculum and the actuality of assessments.

3. Breadth of coverage alignment

The third dimension of assessment refers to the extent the examination tasks cover all the course objectives within each learning area. This analysis focuses on the relationship between the exam content and the total number of specific objectives set in the course. According to the Webb Alignment Model, a minimum of 50% of the curriculum objectives must be addressed in order to achieve acceptable coverage. The results are presented in Table 6.

Table 6.

Breadth of curriculum coverage

Learning Domain	Total Curriculum Objectives	Objectives Addressed in Exam	Coverage Rate (%)	Alignment Achieved
Language Knowledge	19	8	42.11%	No
Language Skills	46	13	28.27%	No

Note: Alignment threshold ≥ 50% of specific objectives addressed.

The findings reveal that the exam does not meet the minimum coverage threshold in either learning domain. In the Language Knowledge domain, only 42.11% of the objectives are represented, while the Language Skills domain achieves a notably lower coverage rate of 28.27%. These results suggest that a significant portion of the curriculum remains unassessed in the current examination structure.

The limited coverage can be explained by several systemic and practical constraints. First, the breadth of the curriculum itself particularly under the competency-based model, is extensive, encompassing diverse subdomains such as phonology, pragmatics, and discourse-level language use. Within the constraints of a single summative assessment, it is often not feasible to reflect the full scope of these goals. Second, some content areas are inherently difficult to operationalize in written tests. For example, phonological competence and pragmatic awareness are more effectively evaluated through forma-

tive assessment or interactive performance tasks, which are rarely included in large-scale written exams. Consequently, objectives related to pronunciation, stress patterns, and language use in social contexts tend to be underrepresented or entirely absent. Furthermore, curriculum goals in the Language Skills domain often require complex, integrated language use, which can be difficult to assess reliably using standardized items. Objectives such as "inferring the meaning of unfamiliar words in context" or "organizing ideas coherently in writing" require open-ended responses or multilayered tasks that are challenging to score consistently, especially in large-scale testing environments.

Qualitative data from teacher interviews support these findings. Interviews with English teachers revealed that many practical competencies, such as expressing opinions in writing or interpreting a speaker's tone in listening tasks, are rarely included in exam design. One of the teachers noted: "Tasks that require students to express subjective opinions or infer implied meanings are often avoided because they are harder to score objectively." This leads to a strategic narrowing of task design, where measurable but low-level competencies are overemphasized.

The failure to achieve an acceptable breadth of coverage poses a challenge to the construct validity of the assessment. It is extremely difficult to make reliable inferences about students' overall language proficiency when important components of the curriculum are not assessed. A narrowing of the assessment scope can also affect teaching practices, leading teachers to overemphasize frequently tested objectives and neglect students' overall language development.

To better align with the anticipated curriculum, on the one hand, it is necessary to expand the assessment framework by incorporating more diverse tasks to examine objectives that are assessed less frequently. On the other hand, it is also possible to comprehensively and evenly reflect future curriculum content by adding technology-enhanced assessments, teacher-assessed components, or integrated performance tasks.

4. Balance of distribution alignment

The final alignment dimension examines whether the exam tasks are evenly distributed across the course objectives. The purpose is to determine whether the assessment gives disproportionate attention to certain goals while neglect-

ing others. Uneven task distribution may compromise both the fairness and representativeness of the assessment. According to the Webb Alignment Model, a balance coefficient of 0.70 or higher indicates an acceptable distribution. The distributional balance for both learning domains is summarized in Table 7.

Table 7.

Distribution Balance of Exam Tasks Across Curriculum Objectives

Learning Domain	Total Curriculum Objectives	Objectives Addressed in Exam	Balance Coefficient (I)	Alignment Achieved
Language Knowledge	19	8	0.70	Yes
Language Skills	46	13	0.54	No

Note: Acceptable distribution threshold = balance coefficient ≥ 0.70.

The analysis reveals a contrasting pattern of distributional balance between the two learning domains. In the Language Knowledge domain, the coefficient reaches the threshold of 0.70, indicating a relatively even spread of tasks across the addressed objectives. This result reflects the broader applicability of vocabulary and grammar items, which can be embedded in multiple task types such as cloze procedures, sentence transformations, or text completion without compromising task reliability or format standardization. Furthermore, the curriculum statements for language knowledge tend to be more general in scope, making it easier to design tasks that align with multiple objectives simultaneously.

In contrast, the Language Skills domain falls short of the acceptable threshold, with a balance coefficient of only 0.54. This implies that tasks in this domain are concentrated around a limited number of objectives, particularly those involving literal comprehension or surface-level information retrieval. As a result, other skill areas such as inferencing, evaluating perspectives, discourse organization, and writing fluency are underrepresented or omitted altogether.

Language skills curriculum objectives are often broken down into very fine details, which presents a significant challenge in task allocation. Many objectives are defined very specifically, such as “write a short invitation,” “infer the speaker’s intention,” or “determine the meaning of a sentence from context.” However, the types of actual exam tasks are relatively limited, making it difficult to effectively assess many specific objectives. This

results in an uneven distribution of tasks in exams. Some tasks only include a few of objectives, but some others are fully ignored.

The frequent reuse of similar question types like listening, reading, and writing section, these further exacerbates the imbalance. For example, reading comprehension often uses formulaic questions that focus on testing students' memory of factual information. But the scope of writing questions is also relatively narrow, failing to cover the various text types and communicative intentions, which is mentioned in the curriculum. Teachers and assessment designers interviewed in this study noted that this lack of task variety stems partly from constraints related to grading reliability and partly from the absence of sufficient exemplars for innovative item types under the reformed curriculum. One of the teachers specifically pointed out the repetitive nature of reading sections: "We've been using the same five-option multiple choice format for years. There's no space for tasks that ask students to evaluate an author's opinion or connect ideas across paragraphs." This over-reliance on predictable formats reinforces the imbalance across skill objectives.

The uneven distribution observed in the Language Skills domain raises concerns about the instructional consequences of assessment design. When only certain objectives are repeatedly assessed, teachers may focus disproportionately on these "tested" competencies, leading to curricular narrowing and reducing opportunities for students to develop higher-level skills. This undermines the broader goals of the reform, which emphasize comprehensive communicative competence, critical thinking, and creative language use.

Improving distributional balance requires more intentional design of assessment blueprints, ensuring that task types correspond to a wider range of curriculum objectives. This may involve introducing more diverse writing prompts, expanding reading item formats to include analysis and interpretation, and piloting performance-based listening tasks. Such measures would better reflect the multidimensional nature of language competence and support more equitable learning outcomes.

The alignment analysis reveals a mixed pattern of correspondence between the 2024 English graduation examination and the 2022 national curriculum standards. Although the exam nearly corresponds with the curriculum objectives in terms of content categories and cognitivity, it still falls short

in coverage and in the balance of task distribution, especially in language skills. This imbalance reflects a disconnection between curriculum goals and the actual exam structure, which affects the exam's ability to comprehensively assess students' integrated language proficiency.

From the perspective of education, these findings highlight the need for a more systematic approach to test design. This approach not only emphasizes validity and reliability but also focuses on representativeness and instructional relevance. Putting curriculum objectives and assessment practice together requires a more balanced integration. This will be essential to support the goal of Chinese competency-based education reform and to ensure fair and comprehensive evaluation of student learning outcomes. Insights from teacher interviews reinforce the quantitative findings and suggest that without targeted professional development and clearer item design guidelines, many local educators remain hesitant to implement assessment innovations aligned with the new curriculum.

Conclusion

This study investigated the degree of alignment between English graduation examination tasks and the Compulsory Education English Curriculum Standards (2022 Edition) in the context of China's ongoing competency-based education reform. Using the Webb Alignment Model as an analytical framework, the research evaluated the correspondence between exam content and curriculum objectives across four dimensions: knowledge category, cognitive complexity, breadth of coverage, and balance of distribution.

The findings indicate a mixed alignment pattern. On the one hand, the exams reached a relatively good level in terms of knowledge categories and cognitive complexity. This was especially true in the language knowledge domain, where tasks clearly reflected the comprehensive language use emphasized in the curriculum. On the other hand, the exams showed clear weaknesses in coverage and task distribution. These issues were particularly notable in the language skills area. Some key learning objectives, such as discourse organization, pragmatic awareness, and reasoning comprehension, were not fully included. In addition, the exam design mostly relied on basic tasks and failed to assess higher-level language abilities effectively.

Qualitative data from teacher interviews also revealed several challenges in the current exam design. These included limited resources, concerns about scoring consistency, and a lack of training in assessment design. These problems affected the application of curriculum principles in real exam applications, so that there is a gap between policy expectations and practice. Concerning his problem, it is necessary to improve support for teachers in the aspect of exam design and strengthen professional development.

In conclusion, while the current examination structure reflects a partial alignment with national standards, particularly in terms of content coverage and cognitive demand, it fails to fully support the broader goals of China's competency-based curriculum reform. For the alignment to be improved, future assessment practices should prioritize broader thematic representation, diversified task types, and the inclusion of higher-order communicative skills. Additionally, systemic support for teacher professional development and assessment literacy is essential to bridge the gap between curriculum and evaluation. These measures will help ensure that final assessments not only reflect what students have learned but also support meaningful, long-term language development aligned with modern educational standards.

Sponsorship information. This research is supported by the 2023 Ministry of Education Youth Fund Project "Research on the Evaluation System for the Quality of Standard Spoken and Written Chinese Language Education in Primary and Secondary Schools in Ethnic Minority Areas of Xinjiang" (Project No. 23XJJC880001).

References

1. Berezina, L. Yu., & Rykova, E. A. (2012). Final assessment of students' academic achievements. *Scientific Research in Education*, *(6)*, 48–52. EDN: https://elibrary.ru/OYIEWH
2. Khutorskoy, A. V., Korol, A. D., & Andrianova, G. A. (2014). Competency-based approach in education: fragments of discussions at the forums of the scientific school. *Eidos*, *(4)*, 19–19. EDN: https://elibrary.ru/UAXWSB
3. Kirst, M. W. (2024). *Standards-Based Education Reforms: Looking Back to Look Forward*. Learning Policy Institute. 100 p. URL: https://eric.ed.gov/?id=ED671010
4. Kovaleva, G. S., & Krasnyanskaya, K. A. (2016). Analysis of assessment systems for the implementation of educational programs in Singapore, South Korea, and China. *Science and School*, *(1)*, 121–131. EDN: https://elibrary.ru/VRGCCB

5. Liu, Jie. (2024). Research on English testing in senior high school graduation exams with a focus on assessing thinking skills (case study of reading and writing). *Advances in Education*, *14*(8), 182. (In Chinese). https://doi.org/10.12677/ae.2024.1481390. EDN: https://elibrary.ru/HUQIAH
6. Maizenger, N. V. (2023). Analysis of the results of the Unified State Exam in English in 2022 in the Altai Territory. *The World of Science, Culture, and Education*, *(1 (98))*, 210–214. https://doi.org/10.24412/1991-5497-2023-198-210-214. EDN: https://elibrary.ru/CNEJGT
7. Malygin, A. A. (2023). Educational assessment: from numeric grades to competency measurement. *Domestic and Foreign Pedagogy*, *2 (92)*, 71–89.
8. Mansurova, S. E., & Tabarovskaya, K. A. (2023). Standardization of general education content and learning outcome requirements: a historical approach. *Man and Education*, *(3 (76))*, 58–66. EDN: https://elibrary.ru/HNHZFM
9. Naydenova, N. N. (2022). Assessment of learning outcomes in the process of standardizing the updated content of general education in foreign countries. *Domestic and Foreign Pedagogy*, *1 (6)*, 140–157. https://doi.org/10.24412/2224-0772-2022-88-140-157. EDN: https://elibrary.ru/VGHXPA
10. Volobueva, T. B. (2021). Designing a web integration model for teacher professional development. *Scientific Support for the System of Personnel Advanced Training*, *(4 (49))*, 31–41. EDN: https://elibrary.ru/QMORCZ
11. Wang, Ya. (2018). Reform of the higher education system in China in the 20th — early 21st centuries. *Society: Sociology, Psychology, Pedagogy*, *(3)*, 62–71. https://doi.org/10.24158/spp.2018.3.12. EDN: https://elibrary.ru/YTUDSM
12. Webb, N. L. (1999). *Alignment of Science and Mathematics Standards and Assessments in Four States* (Research Monograph 18). National Institute for Science Education and Council of Chief State School Officers. 44 p. URL: https://eric.ed.gov/?id=ED440852
13. Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*(1), 7–25. https://doi.org/10.1080/08957340709336728
14. Wenwen, T. (2024). Comparative analysis of the development of foreign language testing in Russia and China. *Moscow Pedagogical Journal*, *(2)*, 85–95. https://doi.org/10.18384/2949-4974-2024-2-85-95. EDN: https://elibrary.ru/UPSHEU
15. Yan, Yinlu, Liu, Min, Liao, Quilin, et al. (2024). Exploring ways to reform modern basic education: a review of the book "Theory and Current Issues of Basic Education Reform". *Advances in Education*, *14*(12), 417. (In Chinese). https://doi.org/10.12677/ae.2024.14122284. EDN: https://elibrary.ru/OHJTGR

Список литературы

1. Березина, Л. Ю., Рыкова, Е. А. (2012). Итоговый контроль учебных достижений обучающихся. *Научные исследования в образовании*, (6), 48–52. EDN: https://elibrary.ru/OYIEWH

2. Хуторской, А. В., Король, А. Д., Андрианова, Г. А. (2014). Компетентностный подход в образовании: фрагменты дискуссий на форумах научной школы. *Эйдос*, (4), 19–19. EDN: https://elibrary.ru/UAXWSB
3. Kirst, M. W. (2024). *Standards-based education reforms: looking back to look forward*. Learning Policy Institute. 100 p. URL: https://eric.ed.gov/?id=ED671010
4. Ковалева, Г. С., Краснянская, К. А. (2016). Анализ систем оценки реализации образовательных программ в Сингапуре, Южной Корее и Китае. *Наука и школа*, (1), 121–131. EDN: https://elibrary.ru/VRGCCB
5. Лю, Цзе. (2024). Исследование тестирования по английскому языку в выпускных экзаменах основной школы с акцентом на оценку мыслительных качеств (на примере чтения и письма). *Advances in Education*, *14*(8), 182. (на кит. яз.). https://doi.org/10.12677/ae.2024.1481390. EDN: https://elibrary.ru/HUQIAH
6. Майзенгер, Н. В. (2023). Анализ результатов единого государственного экзамена по английскому языку в 2022 году в Алтайском крае. *Мир науки, культуры, образования*, (1 (98)), 210–214. https://doi.org/10.24412/1991-5497-2023-198-210-214. EDN: https://elibrary.ru/CNEJGT
7. Малыгин, А. А. (2023). Образовательное оценивание: от отметок в форме цифр до измерения компетенций. *Отечественная и зарубежная педагогика*, *2* (92), 71–89.
8. Мансурова, С. Е., Табаровская, К. А. (2023). Стандартизация содержания общего образования и требований к результатам обучения: исторический подход. *Человек и образование*, (3 (76)), 58–66. EDN: https://elibrary.ru/HNHZFM
9. Найденова, Н. Н. (2022). Оценивание результатов обучения в процессе стандартизации обновлённого содержания общего образования в зарубежных странах. *Отечественная и зарубежная педагогика*, *1* (6), 140–157. https://doi.org/10.24412/2224-0772-2022-88-140-157. EDN: https://elibrary.ru/VGHXPA
10. Волобуева, Т. Б. (2021). Проектирование модели веб-интеграции для повышения квалификации педагогов. *Научное обеспечение системы повышения квалификации кадров*, (4 (49)), 31–41. EDN: https://elibrary.ru/QMORCZ
11. Ван, Я. (2018). Реформирование системы высшего образования в Китае в XX – начале XXI вв. *Общество: социология, психология, педагогика*, (3), 62–71. https://doi.org/10.24158/spp.2018.3.12. EDN: https://elibrary.ru/YTUDSM
12. Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph 18). National Institute for Science Education and Council of Chief State School Officers. 44 p. URL: https://eric.ed.gov/?id=ED440852
13. Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*(1), 7–25. https://doi.org/10.1080/08957340709336728
14. Вэньвэнь, Т. (2024). Сопоставительный анализ развития тестирования по иностранным языкам в России и Китае. *Московский педагогический журнал*, (2),

85–95. https://doi.org/10.18384/2949-4974-2024-2-85-95. EDN: https://elibrary.ru/UPSHEU

15. Янь, Иньлу, Лю, Минь, Ляо, Цюйлинь и др. (2024). Изучение путей реформы современного базового образования: рецензия на книгу «Теория и актуальные проблемы реформы базового образования». *Advances in Education*, *14*(12), 417. (на кит. яз.). https://doi.org/10.12677/ae.2024.14122284. EDN: https://elibrary.ru/OHJTGR

DATA ABOUT THE AUTHORS

Wu Xi, Ph.D. in Pedagogy, Assoc. Prof.
College of Foreign Languages, Xinjiang Normal University
100, Guanjing Road, Shuimogou District, Urumqi, China
434489273@qq.com
ORCID: https://orcid.org/0009-0004-4309-1421

Xie Zhongyuan, Master
College of Foreign Languages, Xinjiang Normal University
100, Guanjing Road, Shuimogou District, Urumqi, China
1181265222@qq.com
ORCID: https://orcid.org/0009-0009-3796-208X

ДАННЫЕ ОБ АВТОРАХ

У Си, кандидат педагогических наук, доцент
Институт иностранных языков, Синьцзянский педагогический университет
ул. Гуаньцзин, 100, район Шуймогу, г. Урумчи, Китай
434489273@qq.com

Се Чжунъюань, магистр
Институт иностранных языков, Синьцзянский педагогический университет
ул. Гуаньцзин, 100, район Шуймогу, г. Урумчи, Китай
1181265222@qq.com

Поступила 18.10.2025
После рецензирования 05.12.2025
Принята 15.12.2025

Received 18.10.2025
Revised 05.12.2025
Accepted 15.12.2025