

Analysis of the Quality of Junior High School English Test Papers Based on Bachman's Framework of Test Usefulness: A Case Study of the 2024 Mid-term Test Paper for Grade 8 in Yichang

Jiarui Xie, Lingling Liu

School of Foreign Languages, China Three Gorges University, Yichang 443000, Hubei, China

Abstract: *With the advancement of educational reform, the importance of English testing in basic education has become increasingly prominent. It is not only an important means of assessing students' learning outcomes but also a key tool for guiding teaching directions. This study is based on Bachman and Palmer's (1996) 'test usefulness', aiming to comprehensively analyse and evaluate the quality of the mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang from the perspectives of reliability, construct validity, authenticity, interactiveness, impact, and practicality. The study hopes to provide some guidance for the design of junior high school English exam papers and English teachers' teaching practices. The research findings indicate that the test paper has high reliability, good impact, and strong practicality. However, there are still some issues with the test paper. First, in terms of construct validity, the test paper comprehensively assesses language ability, cultural awareness, and language knowledge, but it lacks depth in exploring critical thinking; and the test paper does not assess students' learning ability. Second, the test paper lacks authenticity, as the multiple-choice questions are not real-world tasks. Third, the test's interactiveness is weak. Most questions focus on understanding surface-level information in the text, so it limit its ability to activate test-takers' topic-related knowledge.*

Keywords: English language testing, Test usefulness, Test paper.

1. Introduction

With the acceleration of globalization, English plays an increasingly important role in the education system as an important tool for international communication (Crystal, 2003). At the basic education level, English language testing is not only an important way to assess students' learning outcomes, but also a key tool to guide the direction of teaching and learning.

According to the requirements outlined in the *Opinions on Deepening Education and Teaching Reform to Comprehensively Improve the Quality of Compulsory Education* (Central Committee of the Communist Party of China & State Council, 2019) and *Opinions on Strengthening the Proposition Work for Junior High School Academic Proficiency Examinations* (Ministry of Education of the People's Republic of China, 2019), the senior high school examinations should adhere to the principle of testing what is taught, and use curriculum standards as the basis for test design. These assessments emphasize subject core competencies, and focus not only on evaluating students' mastery of foundational knowledge and basic skills, but more importantly on assessing cognitive processes, innovative thinking, and problem-solving abilities. However, there are some problems in the design of English tests in many schools, such as focusing too much on grammar and vocabulary knowledge and neglecting the assessment of students' practical language using ability (Brown, 2004), and the single format of the test, which lacks interactivity and authenticity and is difficult to stimulate students' interest and potential in learning (Weir, 2005b). These problems not only affect the validity of the test, but also pose challenges to the improvement and development of English language teaching. Therefore, as an important means of measuring students'

English proficiency, the quality of the design and implementation of language tests is directly related to the effectiveness of teaching and the enhancement of students' motivation to learn. And how to effectively evaluate the quality of language tests is of vital importance. Bachman and Palmer (1996) suggested that the most important considerations when designing and developing a language test are the purpose of the test and the usefulness of the test, and therefore they established a model for evaluating the usefulness of a test. This model can be used to evaluate the quality of a test and can be expressed by the following formula (Bachman & Palmer, 1996):

$$\text{Usefulness} = \text{Reliability} + \text{Construct validity} + \text{Authenticity} + \text{Interactiveness} + \text{Impact} + \text{Practicality}$$

The formula of test usefulness provides a comprehensive and systematic framework for the design and evaluation of language tests, emphasizing that tests should be considered from multiple dimensions to ensure that they can truly and effectively reflect students' language proficiency and have a positive feedback effect on teaching.

The mid-term examination for Grade 8 English subject in the fall of 2024 in Yichang is a mid-term test organized by some key middle schools in Yichang. These schools are benchmark institutions in the field of basic education in Yichang, and they are consistently ranking at the top of the city's junior high school education. They lead in teaching quality, faculty level, and student advancement rates. These schools, in collaboration with Yichang's English subject researchers, a team of outstanding front-line teachers, and experts, has developed the test paper for this examination. The test's design philosophy, examination direction, and difficulty settings play a significant guiding role in English teaching

within the region. As leading schools for junior high school education in Yichang, their exam design is often emulated by other schools. Therefore, a study based on Bachman and Palmer’s test usefulness to systematically assess the test paper can provide scientific evidence for optimizing test proposition by research and teaching departments. Moreover, the problems revealed in the test can also feedback into teaching practice, and offer direction for teachers to improve their teaching, thereby promoting the overall improvement of English assessment quality in Yichang.

In this study, the mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang is taken as the object of study, and based on Bachman and Palmer’s (1996) principle of “test usefulness”, the test paper is comprehensively analyzed and evaluated from the perspectives of reliability, construct validity, authenticity, interactivity, impact and practicality. The study is not only concerned with the quality improvement of a single set of test papers, but also aims to dissect the “model case” of test questions set by a well-known school. Through in-depth discussion of the paper in various dimensions, we aim to reveal its strengths and weaknesses, provide new directions for improving English teaching and assessment in Yichang, and contribute local insights to the ongoing reform of English examinations in China’s basic education system.

2. Description of the Test Paper

The mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang was written by English teachers with rich teaching experience, and they went through several rounds of review to ensure the scientific, fair and relevant nature of the test paper. The design of the test is based on English Curriculum Standards for Compulsory Education (2022 Edition) and English Curriculum for Chinese Primary Schools and Junior Schools (2011 Edition). The test paper is worth a total of 90 points and consists of two parts: objective questions and subjective questions. Objective questions include three types of questions: cloze, reading comprehension, and short passage completion. The subjective questions are a combination of reading and writing, including task-based reading and writing. There are a total of 51 questions on the exam paper.

The purpose of this test is to comprehensively assess students’ mid-term learning outcomes based on academic quality standards and curriculum content. It aims to find out whether students are able to use what they have learned in English flexibly and whether they are ready to continue learning English at a higher level. Through this test, teachers can understand their strengths and weaknesses and provide suggestions for subsequent teaching.

In terms of the type of test, this test is a diagnostic test used to assess the English learning outcomes of eighth-grade students. In terms of test orientation, this exam is part of a formative for learning test. Since the total score of this exam is 90 points and the passing score is 54 points, this test also falls under the category of a criterion-referenced test.

In terms of the approach of test, the paper is a direct testing. The paper examines both language knowledge and skills such

as vocabulary, grammar and reading, as well as comprehensive language abilities such as, reading comprehension and writing, and therefore the test paper is also a combination of a discrete point test and an integrative test. Moreover, the overall test paper includes both objective and subjective questions, so the test includes both objective and subjective tests. The details are shown in Table 1.

Table 1: Types, orientations & approaches of language tests

<i>Test paper</i>	<i>Type</i>	<i>Orientation</i>	<i>Approach</i>
The mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang	Diagnostic Tests	Formative for learning Tests Criterion-Referenced Tests	Direct Testing Discrete Point Testing Integrative Testing Objective Tests Subjective Tests

This test is for eighth-grade students in Yichang. Grade 8 students are between 13 and 14 years old, in early adolescence, with rapid physical and mental development and strong learning ability. They have received basic English education in primary school and continue to learn English after entering junior high school, and have had at least six years of English learning experience. The textbook that the students are using is the People’s Education Press edition “Go for It” for Grade 8, Volume 1, and they had studied up to Unit 7 before taking this test.

3. Analysis of the Quality of English Test Papers Based on Bachman’s Test Usefulness

3.1 Reliability

According to Bachman and Palmer (1996), reliability refers to the consistency and stability of test results. Specifically, reliability reflects the ability of a test to produce consistent results over time, under different testing conditions, and between different raters. A test with high reliability ensures that its measurements are repeatable, that means if the test is administered multiple times under similar conditions, similar results should be obtained.

Zou Shen (2011) summarized three methods to measure testing reliability: Test-retest Method, Alternate-form Method, and Split-half Method. The first two methods have certain challenges in practical application, such as test paper equivalence and time intervals. To avoid these challenges, this study uses the split-half method to establish test reliability. The split-half method treats an exam paper as two relatively independent and corresponding parts. By calculating the correlation between of each part, the internal consistency of the test can be assessed and then it can used to determine the reliability of the entire exam paper. A higher correlation coefficient indicates greater consistency between the two parts, and it results in higher test reliability (Zou, 2011). This study used the Split-Half Method to calculate the test reliability of test paper, which find out a reliability coefficient of 0.891. According to educational measurement standards, if a reliability coefficient is above 0.8, that means the test has high stability and reliability. So, this result suggests that the test content is scientifically and reasonably designed, and students’ test scores can accurately reflect their learning status.

Based on the students’ scores, the descriptive statistical data

for this set of test papers are presented in Table 2. They are including the mean, mode, median, variance, and standard deviation.

Table 2: Descriptive statistics of the test

Mean	Plurality	Median	Variance	Standard variance
66.13	81	70	230	15.17

The average score (66.13) is close to 73.5% of the full mark (66.13/90), indicating that the overall difficulty of the exam paper is moderate, with most students able to complete the main tasks, and it can demonstrate a certain level of language proficiency. The median score (70) indicates that half of the students scored above 70, while the other half scored below 70, with a relatively balanced distribution. The mode (81) indicates that the highest number of students scored 81, suggesting that the exam effectively reflects the true ability of high-performing students while also possessing the ability to distinguish between high-scoring students. The highest score (86) is close to the full score (90), indicating that the exam has a certain upper limit in distinguishing between outstanding students. The minimum score (17) is significantly lower than the average score (66.13), indicating that the exam paper can effectively reflect the ability gaps of students with weaker foundations. The variance and standard deviation show that the score distribution has a certain degree of dispersion, indicating that there are significant differences among students, and the exam paper can distinguish between students of different levels, with good differentiation. The mean, median, and mode of the scores are reasonably distributed, and combined with the moderate standard deviation (15.17), this further validates the stability and consistency of the exam paper.

3.2 Construct Validity

According to Bachman and Palmer's (1996) test usefulness, construct validity refers to whether a test accurately measures the language proficiency constructs it defines. Specifically, construct validity is concerned with whether a test is a comprehensive and accurate reflection of the language abilities it is intended to measure, and whether these abilities are consistent with the test's design objectives.

Department of education of Hubei Province (2023) states that the construction of English test questions aims to implement the fundamental task of fostering moral integrity and nurturing talent. This is based on the objectives, content, and academic quality standards of the junior high school English curriculum as specified in English Curriculum for Chinese Primary Schools and Junior Schools (2011 Edition), while also reflecting the new concepts introduced in English Curriculum Standards for Compulsory Education (2022 Edition). Therefore, based on the National Curriculum Standards for Compulsory Education, this construct of ability includes language ability, cultural awareness, critical thinking skills, learning ability, language knowledge, and language skills. Language knowledge encompasses phonetics, vocabulary, grammar, discourse, and pragmatics; language skills include receptive skills and productive skills.

Part I of the test paper includes one cloze and four reading comprehension passages, while Part II includes one short passage with gaps, one task-based reading, and a written. The cloze consists of 15 questions, and the passage is a narrative text telling a story about friendship and mutual assistance. It assesses vocabulary understanding (Questions 1, 4, 9), text comprehension (Questions 2, 6, 15), grammar knowledge (Questions 3, 11, 13), and logical reasoning (Questions 5, 7, 14). The Reading Comprehension includes four reading passages, two of which are expository writings, one narrative writing, and one argumentative writing. Text A is an expository writing that provides a travel guide to several famous landmarks in Liverpool. This reading comprehension passage simulates students' ability to obtain and analyse practical information while travelling or consulting a travel guide, testing their information retrieval, synthesis, and simple inference, as well as the ability to understand the main idea of the text. Text B is a narrative writing that tells the story of the growth of renowned British dancer Gillian Lynne. Students need to read about Gillian Lynne's growth experience, analyse the plot and core information of the character, corresponding to the context of reading biographies or inspirational stories, emphasizing the understanding of narrative structure and thematic significance. This reading comprehension passage assesses students' ability to understand narrative writing and analyse character stories. Text C is an argumentative writing on healthy eating habits, which focus on how parents help children to develop healthy eating habits at home. Students need to analyse the content of the text, extract healthy eating recommendations and their significance for families. This reading comprehension test assesses students' ability to extract specific information and apply suggestions. The text of six-in-five cloze Test is an expository writing. It is about how to take good photos, and it focus on several key techniques. Students need to reorganize the paragraphs and describe photography techniques in logical order to ensure coherence between paragraphs. This question can assesses students' ability to understand structure and logic.

Part II includes short passage completion, task-based reading, and written expression. The short passage completion is a narrative writing, which tells a story about family activities. Students must select appropriate vocabulary and fill in the blanks in the correct format. This question can assess students' vocabulary, grammar usage, and mastery of discourse coherence. The task-based reading is a biographical writing about the life of Walt Disney. The written expression requires students to write a letter to their pen pal Dave, and they need to share information about their new friend Nick, the similarities and differences between themselves and Nick.

In this section, the test items in the mid-term English test of Grade 8 were linked to the construct of ability, and the number of construct of ability sub-items was compared to the number of competencies tested in the test paper (see Table 3). This analysis is crucial in determining whether our test is a valid measure of students' English proficiency.

Table 3: Comparison of the number of construct of ability sub-items with the number of competencies measured in the test paper

<i>Construct of Ability</i>	<i>Number of construct of ability sub-items (items)</i>	<i>Number of competencies measured in the test paper (items)</i>	<i>Proportion of the number of competencies measured in the test paper to the number of construct of ability sub-items(%)</i>
Language ability	14	6	43
Cultural awareness	17	7	41
Learning ability	10	0	0
Critical thinking	10	4	40
Vocabulary	5	4	80
Grammar	4	3	75
Discourse	4	4	100
Pragmatics	3	2	67
Comprehensible language skills	13	3	23
Expressive language skills	9	3	33

As can be seen from Table 3, there are 14 items of language ability required to be attained by the Grade 8 section in English Curriculum Standards for Compulsory Education (2022 Edition), and this paper tested six of them, including the three dimensions of perception and accumulation, acquisition and construction, and expression and communication; there are 17 items of cultural awareness required to be attained by the Grade 8 section, and this paper tested seven of them, including the three dimensions of comparison and judgement, adaptation and communication, and perception and internalization; there are 10 items of critical thinking required to be attained by the Grade 8 section, and the paper tested 4 of them, including the dimensions of observation and discernment, induction and inference, but did not assess the dimension of criticism and creativity; there are 10 items of learning ability required to be attained by the Grade 8 section, but this paper did not assess the students' learning abilities.

From the table and analysis above, it can be observed that this text paper places a significant emphasis on language ability (43%). Since this study only analyses the written portion of this examination paper and does not cover the listening and speaking sections, the listening and speaking components of this paper are not analyzed. Excluding the listening and speaking components from the construct of ability, the remaining language abilities are almost all assessed in this exam paper. The test paper also places significant emphasis on cultural awareness (41%), as it is based on discourse as a vehicle, with each text carrying cultural content. This allows the test to assess both students' language ability and their cultural awareness. However, in terms of critical thinking, the test primarily assesses students' observation and analysis, as well as their ability to generalize and reason, without delving into deeper aspects such as critical thinking and innovation. The test paper also barely touches on the measurement of learning abilities.

The test paper also places significant emphasis on testing vocabulary (80%), grammar (75%), text knowledge (100%), and discourse (67%) within language knowledge. Since the test paper is structured around discourses, the proportion of discourse testing is relatively high. So this test paper prioritizes testing vocabulary, grammar, discourse, and pragmatics. Regarding language skills, since most are related to listening and speaking, they are not extensively assessed in the written exam.

To sum up, there is a clear correspondence between the test tasks and the constructs of ability, and most of the language

ability, cultural awareness, language knowledge and language skills are examined, but the examination of critical thinking is still at the surface level, and this test paper does not examine the students' learning ability. The test paper focuses on vocabulary, grammar, discourse and pragmatic knowledge.

3.3 Authenticity

Authenticity refers to the degree to which the characteristics of a particular language testing task correspond to those of target language use tasks (Bachman & Palmer, 1996). They believe that test tasks should simulate real language use situations as much as possible to ensure that the test results can truly reflect the language proficiency of the test takers in real situations. This section analyses the authenticity of the test paper in terms of both the authenticity of the material and the design of the test paper.

3.3.1 Authenticity of Materials

The materials of the paper are all derived from real-life magazines, journals and websites. For example, in the reading comprehension section, text A is from the CNN website, text B is from the article entitled The littler girl who couldn't sit still on Laura Bachynski's website, and text C is from the article entitled Encouraging Healthy Eating Habits In Children: Fun & Practical Tips. In addition, the topics of the materials include friendship and sportsmanship, travelling information, personal growth and education, healthy eating habits, photography skills and family leisure activities, etc. The topics are rich and close to real life and students.

3.3.2 Test Paper Design

The test paper consisted of 51 questions, including 35 multiple-choice questions. There are four types of questions: multiple choice, fill-in-the-blank, short answer and written expression. It can be seen that multiple choice questions accounted for 69% of the total number of questions. But multiple-choice questions are not real tasks. In real life, people rarely face four options and choose one of them to express their understanding (Weir, 1990a). Therefore, the question this test paper questions are designed to be less authentic and the multiple choice questions were not authentic tasks.

3.4 Interactiveness

In Bachman & Palmer's Test Usefulness Principles,

interactiveness is one of the core principles in assessing test usefulness, focusing on the dynamic interaction between the examinee and the test task. Bachman & Palmer, state that interactiveness refers to the degree to which a test task Bachman & Palmer state that interactiveness refers to the extent to which the test task involves and activates the examinee's language ability, topical knowledge and affective schemata (Bachman & Palmer, 1996).

In terms of language ability, the construct validity analysis showed that this test paper was able to examine some of the competencies required in English Curriculum Standards for Compulsory Education (2022 Edition) and English Curriculum for Chinese Primary Schools and Junior Schools (2011 Edition) which were developed based on the background knowledge and competencies of this age group of students after extensive research conducted by the researchers who developed the standards, and therefore, the language knowledge examined in the paper is appropriate for the examinee and is able to activate their language ability. The language knowledge examined in this paper is therefore appropriate for the examinee and will activate their language abilities. This test paper is rich in topics, including travelling, personal growth and healthy living, which are close to students' lives, but most of the questions in the test focus on understanding the surface information of the text and lack in-depth discussion of the topics. For example, in the reading comprehension section, although there are some inferential judgement questions, these questions are mainly based on simple reasoning based on specific information in the article, without leading examinee to think and discuss the topics in more depth, which made the test somewhat limited in activating examinee' knowledge of the topics and unable to fully test candidates' understanding and analytical ability of different topics. In terms of affective schema, the test covers a wide range of topics, but they are all required English Curriculum Standards for Compulsory Education (2022 Edition) and English Curriculum for Chinese Primary Schools and Junior Schools (2011 Edition) and are familiar to candidates, so the test don't produce affective schema that would affect candidates' performance. Cloze and reading comprehension sections of the test are in the form of multiple-choice questions. This type of question tests candidates' knowledge of the language but has some limitations in terms of interactiveness. Candidates only need to choose the answers from the given options without actively using language to express and communicate, so the interactiveness of this test is relatively low.

3.5 Impact

Bachman & Palmer (1996) defined impact (also known as wash-back effect) as the influence of tests on teachers' and students' behaviours, attitudes, and learning outcomes during the teaching process. Impact can be positive or negative. And this section explores the impact on both test-takers and teachers.

3.5.1 Impact on test-takers

This test paper is a mid-term test paper that serves the dual purpose of formative and summative assessment. Test-takers can understand their own English ability according to this

paper, check for gaps and make up for them, so that they can plan their study programmes for the second half of the semester in a more systematic way, and the autonomy of students' English learning can be strengthened. However, the examination may also bring some negative impacts on students. Test-takers who do not perform well may fall into self-denial and feel frustrated; they will be more busy in the training of test-taking skills, and will be increasingly preoccupied with exam technique training and mock exercises, neglecting the cultivation of in-depth understanding and practical application skills.

3.5.2 Impact on teachers

This test paper assesses students' language knowledge, language skills, language ability, cultural awareness, and critical thinking. It provides teachers with clear guidance on teaching objectives. Through the exam feedback, teachers can clearly understand whether students master different language skills and then they can adjust teaching plans and classroom activities accordingly. Additionally, this test serves as an effective evaluation tool to help teachers assess students' learning outcomes. For example, through analyzing students' performance in cloze tests, teachers can understand the depth of their grammar and vocabulary mastery; and through written expression, teachers can evaluate students' language organization skills and logical thinking. According this, teacher can provide targeted learning recommendations to them. However, a negative impact is that teachers may become more focused on exam content in subsequent teaching, and they teach only what is tested.

3.6 Practicality

Practicality refers to the relationship between available resources and the resources required for testing (Bachman & Palmer, 1996). Bachman and Palmer proposed three types of resources: human resources, material resources, and time. In terms of human resources, the main type of questions pf test paper are multiple-choice questions. This type of question has clear scoring criteria and unique answers, so it can be completed through machine or rapid manual grading, thereby teachers will spend less time and energy on marking test papers. This significantly reduces the workload of scorers and improves scoring efficiency. Even for more subjective questions, such as short passage completion and written expression, the test provides relatively clear scoring criteria. This helps teachers score quickly and accurately. In terms of material resources, the equipment requirements are simple. Since the listening and speaking exam is separate from the written exam, only test papers and answer sheets are needed. And this exam is not a large-scale exam like the high school entrance exam, so the venue requirements are not high, and ordinary classrooms are sufficient. In terms of time, the test duration is 100 minutes and the total score of this test paper is 90 points, which is reasonably scheduled. Students have sufficient time to complete all questions,

Therefore, based on the analysis above, the mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang demonstrates strong practicality from the perspectives of human resources, material resources, and time.

4. Conclusion

The quality of the mid-term test paper for Grade 8 English subject in the fall of 2024 at X Middle School in Yichang was analyzed and evaluated from six aspects: reliability, construct validity, authenticity, interactiveness, impact, and practicality. The results indicate that this test paper has high reliability, good impact, and strong practicality. However, the test paper still has some problems.

First, in terms of construct validity, the test paper comprehensively assesses language ability, cultural awareness, and language knowledge, but it lacks critical thinking, particularly higher-order critical and innovative thinking. There is a gap between this and the objectives of our quality education. Quality education emphasizes the development of higher-order thinking skills in students, not just the mastery of basic knowledge. So it is recommended that more open-ended questions and tasks be added to examination papers in the future in order to encourage students to think deeply and express themselves creatively. Additionally, the test paper does not test students' learning abilities. Perhaps it is because learning ability is a multi-dimensional concept, which not only includes the mastery of language knowledge and skills, but also involves many aspects such as learning strategies, independent learning ability, motivation and attitude. Traditional test papers mainly test students' language knowledge and skills, such as grammar, vocabulary, reading comprehension and writing ability, but it may not assess students' learning ability. Moreover, the assessment of learning ability needs to be carried out in a real learning situation, which is usually lacking in traditional test papers. So it is a better option to examine students' learning ability in their regular studies rather than using traditional test papers. Second, the test paper's authenticity is relatively low. Because this test paper mainly consists of multiple-choice questions which do not represent real-world tasks. Therefore, it is recommended that more authentic tasks be added to the exam papers to improve their authenticity. Third, the test's interactiveness is weak. The written expression can reflect the test's interactiveness best. This part requires students to use language knowledge and writing strategies to complete writing tasks. However, other sections, especially multiple-choice questions, only focus on recognition skills rather than actual language usage abilities. Furthermore, most questions in the test focus on understanding surface-level information in the text, but lack in-depth exploration of the topic. So it is recommended that more in-depth questions be added to the test paper to guide students in deeper thinking and discussion.

In summary, this test paper demonstrates good reliability, impact, and practicality, and it can effectively assess students' foundational language knowledge and basic application skills. However, from the perspective of comprehensive language testing quality, this test exam still has room for improvement in construct validity, authenticity, and interactiveness.

References

- [1] Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press.
- [2] Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. Pearson Education.
- [3] Central Committee of the Communist Party of China & State Council. (2019). Opinions on deepening education and teaching reform to comprehensively improve the quality of compulsory education http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/201907/t20190708_389416.html
- [4] Crystal, D. (2003). *English as a Global Language*. Cambridge University Press.
- [5] Department of Education of Hubei Province. (2023, July 1). Detailed Interpretation of Provincial Unified Question Setting for the High School Entrance Examination. Retrieved from http://jyt.hubei.gov.cn/bmdt/gxhptlm/mtjj/202307/t20230703_4730102.shtml
- [6] Ministry of Education of the People's Republic of China. (2019). Opinions on strengthening the proposition work for junior high school academic proficiency examinations. http://www.moe.gov.cn/srcsite/A06/s3321/201911/t20191128_409951.html
- [7] Weir, C. J. (1990a). *Communicative Language Testing*. London: Prentice Hall International (UK) Ltd.
- [8] Weir, C. J. (2005b). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan.
- [9] Zou, S. (2011). *An Introduction to English Language Testing (3rd ed.)*. Beijing: Higher Education Press.