

A Study on the Characteristics of Writing Test Items of English High School Entrance Examinations in China in 2020 and 2021

Lu Yiruo, Qian Xiaofang *

School of Foreign Languages and Literature, Beijing Normal University, Beijing, China

Email address:

604435065@qq.com (Qian Xiaofang)

*Corresponding author

To cite this article:

Lu Yiruo, Qian Xiaofang. A Study on the Characteristics of Writing Test Items of English High School Entrance Examinations in China in 2020 and 2021. *International Journal of Language and Linguistics*. Vol. 10, No. 4, 2022, pp. 239-247. doi: 10.11648/j.ijll.20221004.13

Received: July 3, 2022; **Accepted:** July 24, 2022; **Published:** August 4, 2022

Abstract: Being the most important large-scale and high-stake examination in compulsory education, the senior high school entrance examination (SHSEE) has been the focus of testing research for a long time. They are designed to examine students' language achievement and proficiency according to the requirements of the 2011 curriculum standards for compulsory education but are the same time influenced by the concept of core competencies by the senior high school English curriculum standards in 2018. The writing test, especially the writing test items in SHSEE can reflect students' English proficiency and give washback to future test design and writing teaching. This study carries out an analysis of 102 writing items in SHSEE from all over the country in 2020 and 2021, focusing on their characteristics and the quality of test design. This paper adjusts previous frameworks and follows three dimensions, seven subdimensions to analyze: Test Content (Genre, Topics), Prompt (Form, Length), Test Context Design (Authenticity, Interactivity, Openness). The results show that: Firstly, practical writing remained the most prevalent. Prompt form tends to diversify and the form of outline prompts remains the most popular. Most of the items succeed in designing authentic, interactive items with open space for students to write and create. Most test designs are in accordance with the concept of core competencies of the latest curriculum standards issued this year. The results of the study has implications and provides some guidelines for future writing test design of middle school English.

Keywords: Senior High School Entrance Examination, English Writing Test, Test Design

1. Introduction

Being one of the most important tests at basic education level in China, the Senior High School Entrance Examinations (hereafter SHSEE), the English *Zhongkao*, influences more than 15 million teenagers. It serves dual purposes as both an achievement test of middle school English learning and a selective test for senior high school. It usually takes the form of a 100 or 120 minute written test. The writing section, termed as written expression test, is the final section of the test, accounting for around 12% of the total score. It is designed to comprehensively examine students' writing ability, stated as students being able to draft and revise small essays independently based on prompts, specified as the "five-level graded objective" (for graduation from Grade 9) in the English Curriculum Standards for Compulsory English Education [4],

but they were influenced by the notion of key competencies goal of senior high school curriculum standards released and implemented in some provinces since 2018. The topics not only mirror the English key competencies standards and focus on examining students' emotions and attitudes, but also relate closely to time. In the last decade, SHSEE tests were administered on either provincial (including municipalities) or city (also regional) levels with varying qualities in designs.

Among studies on large-scale and high-stake examinations like SHSEE, validity has remained a fundamental issue, including validity studies on TOEFL, IELTS worldwide, and National English Matriculation Test (NEMT, Chinese *gaokao*), College English Test (CET) Test for English Majors (TEM) in China. Many studies focus on the content validity of NEMT based on some frameworks, including Li's [16] evaluation frameworks of English writing test items. The latter two

frameworks combine validity theory with the socio-cognitive model, and this paradigm has also been modified and used to a growing number of empirical research.

2. Literature Review

2.1. Validity of Writing Test

Validity is defined as information showing the degree to which a test is capable of attaining specific goals. The understanding of validity has evolved progressively, starting with a fairly broad content-based model and gradually incorporating the criteria model, trait and factor models, the constructed model, and concerns about fairness and other repercussions [3, 7]. Achieving a balance between reliability, validity, and feasibility of writing test questions has become a very important task for test designers. The well-established works on the validity of assessing English writing include the frameworks put forward by Cronbach and other researchers [4]. Among them, the argument-based validity method focuses on two steps: (1) formulating claims regarding the suggested interpretation and application of test results as a coherent, interpretative argument; and (2) testing those claims using theoretical and empirical data relating to test performances and scores. The center of validity framework was later shifted to construct validity. Summarizing the benefits of argument-based validation, the IUC model emphasizes that the validation process is never complete: theory evolution, shifts in educational and testing environments, and critiques of current procedures all need a recursive validation process, which means the exploration of better assessing framework will never end.

The most significant attribute of a test is its utility [2]. Any excellent test should have six characteristics: reliability, construct validity, authenticity, interactiveness, impact, and practicality. They do highlight, however, that the two criteria—reliability and validity—are vital and fundamental measuring properties for every effective test. The socio-cognitive model claimed that the test design should cover many dimensions when we measure the validity of writing tests [18, 19, 25]. To better portray the writing process, they modified the definition of validity theory and reduced the characteristics of theoretical validity to two dimensions [10], of which the context validity dimension encompasses the writing test task design, the setting of administration, and safeguards. The evaluation model of writing tasks needs to take into account the response format (genre), purpose, prompt length, and writer-reader relationship.

2.2. Writing Test Design in Large-Scale and High-Stake Examinations

There have been relatively fewer studies on the writing test in secondary schools, compared to those in college entrance examinations or college English tests, especially after the introduction of the New English Curriculum Standard (2011 edition). For example, Ding [6] conducted a study of 150 English written expression test questions of NMET (*Chinese*

Gaokao) from various provinces and cities across China in 2013, using Koo and Gao's analytical dimension [15].

Koo and Gao [15] used 20 writing test items from the 2007 NMET and the independent papers of various provinces and cities and proposed a framework of analysis based on the three major and six minor dimensions. Dong et al. [9] and Yang et al. [26] updated the original model, and proposed an analytical framework focusing on three major aspects of test content, feed-in information, and test context design, with seven specific sub-dimensions. The new model better distinguishes between trial content and feed-in information, and resolves the ambiguity of the previous model in the classification of scenario provision methods. In addition to the mainstream analytical models, Gao et al. [10] proposed a propositional model for writing tests based on the theoretical foundation of Li [16], which structured the rationality of the propositions in terms of three major steps and five subdivision dimensions: scenario, feed-in, and output. The analytical framework [9] was further modified by dividing test questions into three items of test content, test information, and test design with a total of seven subitems. In addition to high-stakes exams like NMET (e.g. [17]), writing test questions in large-scale English exams for students in higher education have also received attention, such as TEM4 and TEM8, CET (e.g., [8, 15]).

International scholars have also selected IELTS, TOEFL, and ESOL exams as the subjects of writing tests. Some studies focus more on social dimensions [14]; for example, Slomp et al. [20] studied the framework for using consequential validity evidence, paying more attention to the validity of the framework itself, particularly to whether it can precisely explain the situation in every region of Canada, given Canada's multiracial and linguistic regional heterogeneity. Furthermore, most international studies are more concerned with the linguistic usability of the test questions themselves and take more consideration of the environmental and humanity factors, such as the examination of critical thinking skills in expository writing.

The studies on English writing in SHSEE are mostly ephemeral studies; for instance, Ding [6] did ephemeral studies for several provinces and regions, analyzing the changes and processes of written expression test questions within a time period. In addition, some scholars have also analyzed writing test items in relation to the requirements of the contemporary Chinese curriculum standards, particularly the core competencies of the English subject, for example the study of the writing test items from a single year of SHSEE in thirteen cities in one province [23].

2.3. Theoretical Framework

This paper adjusts the previous framework [33] and follows three dimensions, seven sub-dimensions to analyze test items: Test Content (Genre, Topics), Prompt (Form, Length), Test Context Design (Authenticity, Interactivity, Openness).

2.3.1. Test Content

A complete writing task or the writing test content consists of function, genre, object, and topics [13]. In SHSEE, the

function includes writing purposes like invitation, apology, and so on. Object means the readers who test takers should address. Genre and topics are defined later. Both object and purpose vary in different contexts and will be discussed later.

It is difficult to define genre because there are three schools of thought trying to illustrate it. This paper chooses to follow Swales' [21] definition in that genre comprises a class of communicative events, in combination with Weigle's [24] discourse mode, which includes genre, rhetorical task, and patterns of exposition. Hence genre in this framework refers to the expected form and communicative function of the written product but also means the traditional discourse modes of narration, description, and so on [18].

Traditional discourse mode can be categorized into exposition, description, narration, and persuasion [18]. In SHSEE, letter writing is the most usual text type and can be categorized as practical writing. Description is uncommon and seldom appears alone in junior middle school writing. Genres in SHSEE tests will be mainly categorized into four types—narration, exposition, persuasion, and practical writing and analyzed mostly through the content of the writing, not the format. For example, some test asks students to reply in letters, but the content to be produced actually takes the form of persuasive writing such as discussing the benefits and shortcomings of over-parenting.

Topics greatly influences test takers' performance. Hence, candidates should perceive task topics as suitable, realistic, reasonably, familiar, and feasible [11]. The writing test items in SHSEE are expected to set suitable writing topics that enable students to fulfill the task. The provision of stimulus texts will ensure equal access to resources and reduces the potential bias of student's internal knowledge [12]. Twenty-four topics are listed in the English Curriculum Standards. Based on previous studies on SHSEE writing, this study finds six topics most commonly tested upon and they basically cover all writing tests at this level: namely personal backgrounds, school life, after class routines (especially with family), society and culture, interpersonal communication, and hygiene and health.

2.3.2. Test Context Design

Test context design can be analyzed from three perspectives—authenticity, interactivity, and openness [26].

Among the six criteria of the framework to evaluate language tests [2], authenticity is considered an important mark of a language test's quality [6]. Authenticity means that the task setting should well reflect situations that testees will encounter in their real life and embody the characteristics of daily language and communicative competence [16]. Mirroring real-life situations can effectively stimulate students' interest and create an immersive language environment. But authenticity is by no means equal to copying everything in real-life communication. The context design should balance the characteristics of the language task in real life and in a language test. Authenticity can be examined from two perspectives, situational authenticity and interactional

authenticity. The present study focuses more on situational authenticity as it is more closely related to context, whereas the interaction between the testees and the task will be examined from the interactivity dimension.

The interactivity of writing test draws increasing attention in communicative language tests. It measures the degree of involvement of learners in cognitive and affective and behavioral domains [1]. Its connotation is expanded and implies that the writing task should be set clear, concerning the supposed writer, supposed reader and the purpose of writing, and it helps students to cultivate reader consciousness [16]. Test takers are supposed to choose suitable style and register to finish the writing task. To better monitor the quality, test designers should be equipped with pragmatic awareness. Under this communicative and interactional scoring framework, writing test measures multiple competencies.

Openness requires that students be given a certain amount of freedom to play while at the same time within certain confines for the reliability for marking, so it should not be too strong or too weak.

3. Research Design

3.1. Research Questions

In order to explore the characteristics of English writing test items, this study is intended to answer this question: What are the characteristics of the writing test tasks in terms of three dimensions: test content, feed-in information, and test context design?

3.2. Research Subject

This paper focuses on writing tasks with a certain length of discourse (at least 50 words), so writing tasks at the word and sentence level or tasks with less than 50 words, such as responses to questions in reading comprehension, are not included. Tasks that test overall writing ability in each paper are studied and 16 items from both the years 2020 and 2021 on provincially unified test levels and 35 items each year from city unified test items (Table 1), with 102 items in total.

Table 1. Amount of Collected Test Items.

Number	Province Unified Test Items	City Unified Test Items
2020	16	35
2021	16	35
Total	32	70

3.3. Procedure

This study combines the English Curriculum Standards for Compulsory Education [3] and the English test instructions for Chinese language examinations in various provinces. It proposes to modify the analytical framework [26] to examine three major aspects and seven specific sub-dimensions in terms of test content, feed-in information and test context design. There are some revisions in this framework in order to fit in the situations of SHSEE (Table 2).

Table 2. Analytical Framework.

Dimensions	Categories	
Test content	Genre	Narration, Exposition, Persuasion, Practical Writing
	Topic	Personal backgrounds, School life, After class routines (especially with family), Society and culture, Interpersonal communication, Hygiene and health
Feed-in Information	Prompt Form	Paragraph, Chart and Figure, Picture, Outline And Other Combinations
	Prompt Length	Redundant, Sufficient, and Too Little
Test Context Design	Authenticity	Stronger, Weaker
	Interactivity	Stronger, Weaker
	Openness	To Strong, Sufficient, Too Weak

Based on the framework, the 102 test items were statistically processed through Excel, and the researchers conducted a cross-sectional and bi-directional comparative analysis. The results were cross checked and confirmed by two research assistants.

4. Results and Discussion

After detailed analysis, the results from three dimensions following the same structure as the framework are presented as follows.

4.1. Test Content

According to the English Curriculum Standards for

Compulsory Education [3], students of ninth grade are supposed to draft an essay or short message independently, describe people or events and write down operating instructions based on verbal or graphic prompts.

4.1.1. Response Genre

This study found that among the four types of writing, practical writing remained the most prevalent (Table 3), with 45% of tests requiring students to employ written formats such as letters, speeches, and commentaries. However, letters are no longer the most common format, with test designers preferring speech and essay formats to allow students to express themselves without being constrained by the letter style.

Table 3. Writing Response Genre Analysis.

Genre	Practical Writing	Persuasion	Narration	Exposition
Rate	0.45	0.34	0.16	0.12

The proportion of persuasion does not differ much from the proportion of practical writing, indicating that SHSEE writing tests in 2020 and 2021 have high demand on students' ability of critical thinking, not just on communicative functions like apologies and suggestions.

Writing responses of persuasion requires students to think critically about issues and to clearly present viewpoints. It is also found that persuasion type accounts for a larger proportion in provincial/municipality tests than the city SHSEE tests (Table 4).

Table 4. Comparison of Response Genre between Provincial and City Tests.

Genre	Practical Writing	Persuasion	Narration	Exposition
Province /Municipality	0.38	0.50	0.22	0.09
City	0.49	0.27	0.13	0.13

Students in some premier municipalities have a good command of English on average and their test response genre was persuasion. For example, students were asked to explicate their understanding of change and adaption to changes, either in natural and social environment or personal life and study by means of elaborating on one change in their experience and their gains from it. Persuasion was chosen to encourage students to speak out their own opinions, allowing them to explore and expand.

Instead of providing translation like prompts with exact details, this kind of response genre tests require students to think independently and critically and help develop these skills in teaching as backwash effect.

4.1.2. Topics

The results of writing topics are presented in Table 5. The 'Others' categories include topics like people and language learning.

Table 5. Writing Test Topics Analysis.

Topics	School life	After class routines	Society and Culture	Interpersonal communication	Hygiene and Health	Personal Backgrounds	Others
Percentage	17%	33%	10%	5%	12%	6%	20%

The topics of writing cover a wide range of subjects, not only those related to school, but also those related to life

outside classroom. School life and extracurricular activities accounted for 33% and 17% respectively. School-related

themes are close to students' life, and students can use real-life experiences to express their ideas. These themes are ubiquitous, and most candidates are familiar with them, thus there will be no bias in topics that favor certain candidates over others [9]. The writing tests of F-Province in 2020 and 2021 both focused on school life, namely the school's "English Day" events and traditional Chinese medicine activities on campus. At the same time, off-campus activities have received attention from test writers. Students' behavior at home and in the community were also examined, such as family relationship, after-class volunteering.

The topic mirrors the English key competencies goals and focuses on examining students' emotions, attitudes, and values. 10% of topics are closely related to society and culture, especially Chinese traditional culture. The new English curriculums standards in 2018 point out that learning English assists students in better comprehending the world and gaining advanced scientific and cultural information, as well as in transmitting Chinese culture and improving mutual communication and understanding with young people worldwide. For instance, the items of B-Municipality in 2021 thematize traditional cultural treasures and relate to local cultural symbols, which gives students a greater sense of belonging. Many SHSEE testing topics help develop adolescents' cultural identity and love for their hometown and country.

Some topics also highlight the historical context, inviting students to present personal perspectives during the pandemic, and connect individuals with society. For backwash effect, such subjects are useful for encouraging students to pay attention to society and current events in their regular study. 12% of the topics are related to the theme "Hygiene and Health", while 6% of all are related to the

pandemic. COVID-19 has greatly affected people's lifestyles, as well as the daily habits of students. Item of Y-City, H-Province in 2020 invites students to introduce changes in dining habits during the pandemic and in the item of C-City, S-Province students are assumed to be presenters to give insights from their online learning and share protective measure g during lock-downs. Bringing students closer to the times and contextualizing real-life scenarios of language use are also in accordance with the nation's philosophy of education.

4.2. Prompt

Prompt, or feed-in information, is the material provided for the test taker to stimulate their answers [5]. The input of the writing test appears in different types and lengths. As for the background, the impact of prompt on the quality of L2 written production was investigated by [18]. His idea first shows that too much input before writing does nothing good to students' performance. Based on this idea [24] further studies the impact of text length and suggest that writers may rely on the provided text and even translate the known materials into the target language. Thus, the prompt length should be evaluated by test designers and the form of prompt should be clear and concise to avoid ambiguity.

4.2.1. Form of Prompt

Writing test designers usually choose paragraphs, charts and figures, pictures, and outlines to provide necessary resources as prompts. Testee needs to grab key points from prompts, transforming information into a complete form of writing. Different combinations of forms are frequently created for diversification. According to Figure 1, there are six combinations of forms in all 102 writing tests.

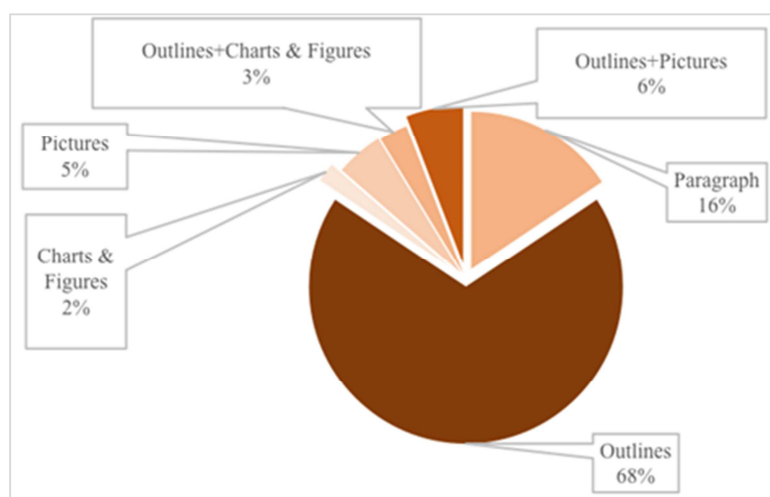


Figure 1. Test Prompt Form Analysis in 2020 and 2021.

Outlines are the form that is highly rated. It weighs close to 70%, far beyond other forms like charts & figures (2%) and pictures (5%). The outline form is succinct and clear and is also favorable for candidates to better understand the writing task. Key points are listed directly in the prompt, which not only assists the examinee in developing a clear composition

framework but also aids in reminding the writer to stay on topic from beginning to end. For example, items of Y-City of H-Province in 2021 is a good example of using outline prompts.

(In Chinese) Suppose you are Li Hua from an international school. Your school is recruiting performers and

volunteers for the annual art festival. Please write a letter to Mr. Smith, the man in charge, to introduce yourself. You also make a few proposals. The letter includes:

- 1) Willingness and reasons;
- 2) Specific plans;
- 3) Reasonable proposals.

(Y-City of H-Province, 2021).

This test task uses outline to succinctly present three main points. Test takers can follow the structure and brainstorm based on the suggested logical order. With 3 points, 13 Chinese characters, students can understand the demand of the test and structure their thoughts. The outline format facilitates analytical marking. One demerit of the well-established outline paradigm may lie in limiting students' creativity to some extent. Therefore, some test designers incorporated charts and pictures to enhance the way information is presented and to assess students' ability to evaluate and synthesize information in multiple ways, which is in line with the cultivation of English key competencies. For example, the writing test of H-Province in 2020 utilized four cartoon graphics and four matching sentences to convey the imagery of medical workers being separated from their families during the pandemic. This test assesses students' observation and analytical abilities by presenting the story sequence and characters in a visual format and utilizing keyword suggestions to aid students in their writing. Similarly, before beginning to write, students must comprehend the horizontal and vertical coordinates of the bar graph and illustrations of pie chart in order to write comprehensively in some other forms. Because combining numerous feed-in information requires students to absorb multiple forms of information at the same time is difficult, fewer tests (16%) employ this format.

4.2.2. Prompt Length

Evidence shows that when reading is incorporated into writing tasks, markers struggle with deciding how much testees can borrow from these sources. If the prompt is in the native language, writing test designs should prevent testees from directly translating the original text [16]. As a consequence, the length of the prompt should be sufficient but avoid redundancy or being too little. The categories of the prompt length include three types: redundant, sufficient, and too little.

Table 6. Test Prompt Form Analysis in 2020 and 2021.

Categories	Amount	Percentage
Redundant	14	14%
Sufficient	72	71%
Too little	16	16%

As shown in Table 6, approximately 71% of the SHSEE writing tests in 2020 and 2021 provide a sufficient degree of prompting information to ensure that test takers were provided enough information but with adequate space for writing freely. "Redundant" and "too little" prompt types account for 14% and 16% respectively. Among these, the writing test of X-Province in 2020 gave test takers far too much material,

allowing the possibility of straightforward translation of the four points.

(*In Chinese*) Good habits benefit our whole life. How to form good habits? Please write a short essay with the title "How To Develop A Good Habit" based on the following prompts and your own thoughts.

Points:

- 1) Keep exercising, don't stay overnight, have enough sleep;
- 2) Eat enough fruit and vegetables and less junk food;
- 3) Study hard, read extensively;
- 4) Love your parents and help them do chores.

(X-Province, 2020).

This writing test contains 122 Chinese characters, half of which can be directly translated, similar to three other writing items in SHSEE in 2020 and 2021. In such tests, the capacity to assess students' thinking will be harmed in such uncritical translation-like writing.

Table 7. Prompt Length Analysis between Provincial and City Tests.

Categories	Province	City
Redundant	16%	10%
Sufficient	66%	75%
Too little	16%	16%

Provincial unified test items provide slightly more redundant prompt than city test items (Table 7). Some provinces may have been required to adjust the degree of difficulty due to educational levels in various regions. City test items outperform provincial test items in providing sufficient prompt.

4.3. Test Context Design

The setting of the writing task in a language test can be divided into three parts—context, input, and output. The important role of context as a determinant of communicative language ability is paramount [18]. Context provides students with real-life background in examinations to utilize the medium 'English' and students can relate to their own experience to solve problems or express themselves in simulated target situations. The quality of test context design influences test takers' performance and can be evaluated in authenticity, interactivity and openness. Authenticity is an essential precondition for the creation of context and guides testees to play the leading role and speak out freely based on their real-life experiences. Interactivity stresses the writer-reader relationship of writing and audience awareness in communication. Writing tests should also give students space for free thoughts and expression, especially in the large-scale high-stakes examination [22]. Request for openness enables students to release their imagination and potential.

4.3.1. Authenticity

Material, response format, and task are chosen as the criteria to evaluate authenticity. This study found that around 70% of the test items fit the requirement of authenticity, but around 30% are weak in authenticity - the context does not

convey the communicative purpose for the situation. For example, if we compare two items that share similar designs with the situations assumed as inviting a foreign friend or an international teacher to campus activities on traditional Chinese medicine or Chinese tea culture, we can find the one from F-Province in 2020 presented more authentic contextual design than the one from J-City, S-Province. Testees have to take the writerly role either as a friend in the former or the chair of student union in the latter. The situation and role dimensions have some degree of plausibility in both, but the factor affecting authenticity is the form of communication, the former being writing an email and the latter writing a letter. Nowadays, students seldom have chances to write formal letters, the outdated form in their perception.

A similar example of low authenticity is the design of writing a journal entry to record a special online PE class held by a world champion in the writing test of G-Province in 2021. Journal entries tend to be personal and thus appear a little strange if its main content is already designated and middle school students rarely use English to record personal experiences in journals. Another example of weak authenticity by communication channels is the test of H-City, H-Province in 2020, in which students have to write a letter to introduce themselves after reading about someone's social media and becoming interested in the activities. In this case, the more frequently used means of communication would be electronic such as comment-and-reply messages, emails, phone messages or even phone calls.

4.3.2. Interactivity

In order to evaluate the interactivity, each writing test of SHSEE was marked as either strong or weak. If the writing test indicates the supposed writer, supposed reader, and writing purpose clearly, it is regarded as strong in interactivity. This study finds that over 80% of items can fulfill the demand for interactivity. Test items of weak interactivity tend to be more like writing essays to state viewpoints like "How to be a better man" or "A question I asked" with no specific genre or readership. Such designs affect scorer and test reliability.

4.3.3. Openness

The study found that context openness in many tests was too strong (Table 8). 27% of items gave students too much imaginative space to write, and 15% of the items showed weak openness. Only 58% of all items provide students with a comfortable and reasonable space to write. In previous studies, most scholars have emphasized that the design of test questions should be more open, so that students can better express their views on a certain topic. Tests should avoid giving too much information laden prompts or answering restrictions, and avoid setting too detailed topic angles. But this study found that many of the tests simply set up a topic around which students started writing directly, which resembles native language writing style, but in large scale high-stakes tests, this kind of openness affects reliability and fairness.

Table 8. Test Openness Analysis.

Categories	Amount	Rate
Too Strong	28	27%
Too Weak	15	15%
Balanced	59	58%

The strength of openness is intrinsically linked to the amount of the prompt. For example, the writing test of A-Province requested students to compose a self-promotion letter to the newspaper, but in the bullet points, they highlighted all of the characteristics of this hypothetical writer. In a case where the presumed writer is generally the student himself or herself for a self-introduction, it is overly restrictive to give the presumed writer a false identity. When students lack the opportunity to introduce themselves, it is difficult for them to express themselves and develop creativity, viewed from the backwash effect perspective.

(*In Chinese*) Suppose you're Li Hua. An English Newspaper is recruiting young journalists from students for their column "My hometown". Please write a letter to introduce yourself. The main points include: 1. Your familiarity with hometown; 2. Good communication skills; 3. Good English skills.

(A-Province, 2021).

On the opposite, overly open-ended questions give students a lot of room to play, but the reliability of the test will be greatly affected when there is no definite direction in terms of topic, purposes and expression. An exemplary test that achieved well-balanced openness and control, however, is the one of S-Municipality in 2021, in which students are oriented to share experiences with the theme of appreciating the uniqueness of everyone with the right amount of guidance and restriction. This kind of items are conducive to developing students' language ability, thinking capacity and cultural awareness, in accordance with the requirement of core competencies in the new curriculum standards for senior high school in 2018 and later the one for junior high school in 2022.

(*In Chinese*) The growing-up experience of Sculpture Williard tells us: No one is perfect, but everyone is special. In Williard's homepage there is a Comment Section and every viewer can leave their story and share their ideas. Please share your experience or what you have heard in the this section, under the title "Everyone can be special".

(S-Municipality, 2021).

5. Conclusion

This study intends to explore the characteristics of writing test items selected from 2020 and 2021 SHSEE. Using the framework of seven subdimensions, 102 items were analyzed. The main findings are summarized. First, the SHSEE writing test items feature a diverse set of topics that cover the majority of students' life experiences and effectively expand them, with the focus on family and school, but not being repetitious or clichéd. Secondly, the types of information provided in the prompt are diverse, with mixed arrangement of texts, photographs, pictures, and charts. The test prompt form is

multimodel, which can not only avoid ambiguity of understanding caused by single text form, but also develop students' comprehensive abilities. This better simulates real-life scenes, and increase the authenticity of the test. Thirdly, in terms of context design, most of the items have fulfilled the requirements of interactivity, which can ensure the writing test context and allow students to communicate with a clear purpose and with the right readerships. The unity of authenticity, interactivity and openness needs to be taken into consideration in those test designs. There is no obvious general pattern of advantages of provincial unified tests over city test items, or vice versa.

There are some implications of the study. First of all, test designers should enrich the topic resources, prompt form and response format. SHSEE writing test should maintain a good balance of practical writing and other types of writing, and include more prompt material like pictures, dialogues, graphs, and avoid "formulaic" writing. Test designers can even create new items that require students to use a combination of different genres to answer. The second point is that test writers should manage the prompt design and the length in particular. It is good to be innovative and allow students more freedom in expression but being overly open-ended in their content while providing too much useless information will disrupts candidates' thoughts and cause the main content of the writing to be obscured. Achieving a balance is important. In order to achieve high authenticity test designers need to adopt the student perspective.

One limitation of the study is that although it contains a sample of 102 English writing test items, it does not represent the characteristics of all English writing tests across the country. And the descriptive analysis is a little simplistic. This study only analyzes the characteristics of writing tests from the perspective of test task design based on a analytical framework, not considering psychological factors from the perspective of students, or being supported with official statistical data. We hope future studies can overcome these shortcomings and evaluate test design from multiple perspectives.

References

- [1] Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: University Press.
- [2] Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: University Press.
- [3] Chinese Ministry of Education (CMOE) (2011) *English Curriculum Standards for Compulsory Education*. Beijing: Beijing Normal University Publishing House.
- [4] Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52 (4), 281–302. American Psychological Association.
- [5] Davies, A. (1999). *Dictionary of language testing*. Studies in language testing. Cambridge, UK: Cambridge University Press.
- [6] Ding, J. (2015). A Study of Written English Expression Questions in the National High School Entrance Examination in 2013. *Journal of Basic English Education*, 17 (01), 105–112.
- [7] Dobrić, N. (2018). Reliability, Validity, and Writing Assessment: A Timeline. *ELOPE*, 15 (2), 9–24. University of Ljubljana, Faculty of Arts.
- [8] Dong M., Fang, X. (2014) The Task Characteristics and Design Principles of TEM4 Writing Test Items--A Diachronic Analysis of TEM4 Writing Test Items From 1992 to 2013. *Foreign Language Testing and Teaching*, (3), 1-10.
- [9] Dong M., Gao X., & Yang Z. (2011). A Diachronic Study of English Writing items in the College Entrance Examination of Nation Unified from 1989 to 2011. *Educational Measurement and Evaluation*, (10), 47-52.
- [10] Gao, X., Li, Z., Zhang, Y. & Huang, P. (2012) How To Design English Writing Test Questions--Take Questions of College Entrance Examination as Examples. 33 (1), 254-256.
- [11] Hamp-Lyons, L. (1990). Second language writing: Assessment issues. *Second language writing: Research insights for the classroom*, 67-68.
- [12] Horowitz, D. (1991). ESL writing assessments: Contradictions and resolutions. *Assessing Second Language Writing in Academic Contexts*, 71–85.
- [13] Hughes, A. (1989). *Testing for language teachers*. Cambridge: University Press.
- [14] Kane, T. S. (2000). *The Oxford Essential Guide to Writing*. Oxford University Press Inc. Large-Scale Assessment Programs for All Students: Validity, Technical Adequacy, and Implementation.
- [15] Koo, X., & Gao, X. (2007) A Synchronic Study of the English Written Test Items in National-based College Entrance Examination and Province-based Autonomous College Entrance Examination Designing in 2007. *Examination Research*, (12), 28-33+36.
- [16] Li, X. (1997). *The science and art of language testing*. Changsha: Hunan Education Publishing House.
- [17] Qi, L. (2006). The Design and Language Application Principle of the English Writing Items in University Entrance Examination. *Foreign Language Teaching in Schools*, 29 (02), 15-18.
- [18] Shaw, S. D., & Weir, S. C. (2019). *Examining writing: Research and practice in assessing second language writing*. Beijing, Foreign Language Teaching and Research Press.
- [19] Slomp, D. (2005). Teaching and assessing language skills: Defining the knowledge that matters. *English Teaching*, 4 (3). University of Waikato, Department of English.
- [20] Slomp, D. H., Corrigan, J. A., & Sugimoto, T. (2014). A Framework for Using Consequential Validity Evidence in Evaluating Large-Scale Writing Assessments: A Canadian Study. *Research in the Teaching of English*, 48 (3), 276–302. National Council of Teachers of English.
- [21] Swales, J. (1990). *Genre analysis: English in academic and research settings*. The Cambridge applied linguistics series. Cambridge: University Press.
- [22] Tao, B. (2010). A Synchronic Study of English Writing Test Items in the College Entrance Examination of 2009. *High School Education: Foreign Language Teaching and Learning*, (03), 58-63.

- [23] Wang, D. & Bian, Y. (2019) A Synchronic Study and Illusions of English Writing Test Items of Cities and Areas in Jiang Su Province in 2018. *English Teacher*, 19 (08), 114-118.
- [24] Weigle, S. C. (2011). *Assessing writing*. Beijing, Foreign Language Teaching and Research Press.
- [25] Weir, C. J. (2005). *Language Testing and Validation, An Evidence-Based Approach*. Palgrave Macmillan UK.
- [26] Yang, Z., Koo X., & Wang, X. (2018) A Diachronic Study of English Writing items in the College Entrance Examination from 2008 to 2017. *Educational Measurement and Evaluation*, (02), 18-26.