

A Research on the Content Validity of the High School Achievement Test Based on the China's Standards of English Language Ability

Qinxuan Li, Hongrui Li, Kangzhou Peng
Chongqing Normal University, Chongqing, China

Abstract

The China's Standards of English Language Ability is a current scale for the ability to guide all levels of education in China. For High School Achievement Test, it is not only necessary to refer to the curriculum standards, but also to develop test tools based on the scale. This study takes the High School Achievement Test as an example, benchmarks the Chinese English Ability Level Scale, analyzes the consistency between the test content and the scale, and provides reference suggestions for the scale to be used in the High School Achievement Test.

Keywords

The China's Standards of English Language Ability; The High School Achievement Test; The Content Validity.

1. Introduction

The "Implementation Opinions on Deepening the Reform of the Examination Enrollment System" issued by the State Council on September 3, 2014 clearly stated that one of the main tasks of strengthening the construction of the foreign language proficiency assessment system is to develop a national foreign language proficiency level scale to provide different levels of foreign language proficiency Standard (Liu Jianda, 2015) [1]. This will help solve the problems of different English test standards in China, separation of teaching and assessment goals, and incoherence of teaching goals at various stages, and realize the "through-train" of English teaching and the communication and mutual recognition of multiple learning outcomes (Lin Huiqing, 2016) [2]. On this basis, the "China's Standard English Language Ability" was officially promulgated in 2018. With the promulgation of the scale, related researches have also emerged one after another. This research will be combined with related research results at home and abroad, and will be reviewed, and it is expected to be helpful to follow-up research.

2. A Brief Review of Foreign Language Examination Research Based on Scales

2.1. Research Status of Foreign Language Proficiency Scale

The "Common European Framework of Reference for Languages: Learning, Teaching, and Evaluation" (CEFR) was first published by the European Commission in 2001 (European Commission [CE], 2001) [3]. Today, European scientists and practitioners are still working in two fundamentally important directions: to improve the descriptive descriptors of the common reference level of each language proficiency, and the results achieved in the practice of language teaching. For example, "Embedding CEFR in the academic field: language task assessment" (Lowiea, Hainesa, and Jansmaa, 2010) [4], "English General Studies 1. L2 Standard Features of English: Specifying the Reference Level of the Common European Framework. 2. Language Functions Interview: Theoretical and empirical basis for the definition of language structure

across competency ranges" (Swan, 2014)[5], "Proficiency" (Harsch, 2017)[6]. Explore the relationship between CEFR levels (A1-C2) and the language of experiential learners (Byram, 2012) [7]. According to the standards described by CEFR, analyze the impact on teachers' teaching (Wu, J, 2012) [8] and so on.

2.2. Current situation of China's Standards of English Language Ability

Since the release of the China's Standard English Language Ability on April 12, 2018, its research has gradually shifted from overall construction to empirical research. After sorting out, it is found that the current research is mainly divided into three aspects: the content of the scale and its validity verification; the docking of the scale with the examination; the practical application of the scale in university teaching. Liu Jianda pointed out that since its implementation, the scale has played an increasingly important role in language test design, curriculum standards, and testing standards formulation (Liu Jianda, 2021) [9]. However, the verification of the validity of the scale is conducive to the better application of the scale in practice. The overall design structure of the scale has been discussed by experts (He Lianzhen and Chen Dajian, 2017) [10]. Based on the development of scales in Europe and other foreign countries, the scale research in China mainly focuses on the principles of scale formulation. The research methods and procedures of the scale. Examination research can be divided into six aspects: listening, reading, translation, writing, interpreting, and pragmatics. Current research is more concerned with college English education. Such as the study of the relationship between listening ability and interpreting ability of translation majors, and the practice of English teaching for physical education students (Mu Lei, Zhang Rong, Chen Guangjiao, 2021) [11]. There are relatively few benchmarking studies on primary and secondary school examinations, or they are concentrated in a certain section (Wu Zunmin, 2019) [12].

2.3. Research Status of Academic Achievement Test

In the database, this study entered the key words "student academic achievement evaluation", "academic achievement" and "high school academic achievement test" respectively, and found that there is relatively little research on academic achievement test in China. And the main research angles of domestic scholars are as follows: firstly, comparative research on achievement evaluation of international mainstream curriculum systems. This research angle mainly introduces the models of academic achievement evaluation in developed countries and provides reference for academic achievement evaluation in China. For example, Comparative Study of Academic Achievement Evaluation Based on Curriculum Standards written by Wang Xianze (2008)[13] analyzes in detail the academic achievement evaluation based on curriculum standards in the United States, Britain, Australia and Canada and its enlightenment to China; secondly, theoretical research of standard-based evaluation. For example, from the historical analysis of the development of students' academic evaluation, Chen Ruisheng (2009) also put forward new thoughts on academic evaluation based on curriculum standards, introduced that the most important feature of academic achievement evaluation based on standards is to establish evaluation according to curriculum standards, and introduced the procedure of evaluation based on curriculum standards[14]. Cui Yunkuo and Xia Xuemei(2007) analyzed why students' academic achievement evaluation should be based on curriculum standards, and what is the significance? How can students' academic achievement evaluation be based on curriculum standards [15]? This is also the theoretical support for us to implement standard-based evaluation. Wang Lei (2007) analyzed the value of academic evaluation-measurement theory under the background of new curriculum, and thought that psychological and educational measurement theory should be applied to academic evaluation. The focus is mainly on how to evaluate based on curriculum standards, that is, the procedure of academic achievement evaluation based on curriculum standards is studied[16]. To sum up, since the initiation of the new curriculum reform, the theoretical research on standards-based academic

achievement evaluation in China has been gradually enriched, but most of them are discussed and studied from the upper theory, and some enlightenment and reference for academic evaluation in China are put forward. How to evaluate students' academic performance based on curriculum standards and how to better evaluate students' academic achievements are rarely discussed in the context of a separate discipline.

Based on the literature review, this study found that the research on the China's Standards of English Language Ability in China is becoming more and more mature. But there is still some deficiency in the process of marking the scale and testing the academic achievement of senior high school. There are few related research documents. So this part will be discussed in the hope of helping to better apply the scale to high school achievement test.

3. Research Design

3.1. Research Questions

The China's Standard English Language Ability and the High School Achievement Test are closely related. The study of the consistency between the test paper and the scale is helpful to improve the content validity of the test questions, thereby cultivating talents that meet the needs of national development to a greater extent. Therefore, the primary questions of this study are: (1) To what extent does the test fit the China's Standard English Language Ability? (2) To what extent is the test compatible with the scale?

3.2. Research Tools

(1) The English test of the 2021 high school midterm exam in a middle school. It has four parts, with a maximum score of 150 points. The first part of listening comprehension consists of two sections with a full score of 20; the second part of reading comprehension has a full score of 37.5, and the second section selects five from seven, with a full score of 12.5; the third part uses language in the first section of cloze, with a full score of 15, The second section of grammar fill in the blanks, a full score of 15; the fourth part of writing the first section of word spelling, a full score of 10, the second section according to the prompts to translate sentences, a full score of 10, and the third section of written expressions, a total of 20 points.

(2) Analysis tools. This study uses the China's Standards of English Language Ability to analyze the test questions

3.3. Research Process

(1) A comparative analysis of the degree of agreement between the overall test questions and the China's Standards of English Language Ability. A detailed comparison of it to define the skills to be examined in the four parts of listening, reading, language use, and writing. After multiple rounds of discussion, the skill attributes of each question were determined. Count the proportions of test questions at different levels of ability and draw the question that the test questions do not achieve better content validity.

(2) Based on the results of the comparison, propose to modify sample questions for the writing test questions to make them more in line with the requirements of the Chinese English Proficiency Scale. First, assess the degree of agreement between the test questions and the scale. Then, select the writing test questions and adapt the existing problems in the writing test questions in combination with the four-level writing ability requirements of the general scale.

3.4. Results and Discussion

3.4.1. The Degree of Consistency Between English Test Questions of A Middle School and the Existing Standards

Appendix 1 presents the results of analyzing the China's Standards of English Language Ability. The overall test questions in the listening part are basically in line with the ability level of the

scale. At the same time, the results reflect two problems of the listening test questions: 1. The skills specified in the scale cannot directly correspond to the test questions; 2. The third and fourth levels of the scale require diversified listening genres, but the listening genre in the test questions does not meet the requirements. First of all, the first grade test question corresponds to the second level of the listening comprehension abilities scale, and the key skill of the listening requirement test is to understand the speaker's point of view and intention, instead of focusing on the recognition and understanding of the details. That is, the proposition of the test question and the level and requirements of the scale aren't accurate alignment. At the same time, the third and fourth levels of the scale require the diversity of listening genres, such as news reports, speeches, announcements, etc., but the listening materials in this set of test questions are almost all dialogues. If the questions are set in this way, it is bound to influence the content validity of the set of test. Even if the student's performance data is obtained, the teacher will still have a deviation when analyzing whether the student has the skills specified by the scale.

The overall reading test questions are consistent with the level of ability of the scale, but the five parts selected from seven are consistent. At the same time, the results reflect that the reading test questions present two problems: 1. The coverage of the survey points is not comprehensive; 2. The cognitive dimension survey level is mostly low-level ability. First of all, the reading questions cover 5 skills related to the grade scale as a whole. Among them, the fourth-level ability accounts for 40%, and the third-level ability accounts for 60%. The sub-table comparison result is contrary to the general table, and the four-level ability survey accounts for 60%. Analyzing the comparison table is not difficult to find that the actual investigation points of the topic are not rich enough, such as 28 and 29 questions to investigate and read stories to identify specific events.

The whole part of the language knowledge conforms to ability level of the scale. At the same time, the results reflect the two problems of cloze and grammar: 1. Some of the cloze questions did not use the ability scale as the standard; 2. The questions in this part did not fully examine the language knowledge points; First, the 48, 52, and 53 questions in the form fill-in-the-blank part cannot find the corresponding standard in the scale. Secondly, according to the four-level requirements of the General Form of Organizational Ability in the scale, the language knowledge points examined in this part are not comprehensive enough. For example, in the cloze and grammar parts, there are no questions which needs students to use grammar such as tense and voice in order to accurately understand the topic of information. The content validity of this set of test questions will inevitably be affected.

The major part of written expression is consistent with the scale ability level. However, the spelling part of the words in the first section of writing and the sentences translated according to the prompts in the second section are not partially consistent. At the same time, the results show that there are three problems in the writing test. First, the lack of authenticity. Secondly, the cognitive dimension is single, all of which are identification. Finally, the spelling of words in the first section of the writing section and the sentences translated according to the prompts in the second section should not belong to the writing section.

3.4.2. Compatibility between the Abilities Examined in the Test Questions and the Abilities Specified in the Scale

First of all, the test questions in the listening part are not difficult. The ability to be investigated is relatively single, and the level of difficulty is not reached. Combined with cognitive skills, the test questions cover two types of low-level skills, recognition and understanding. Recognition is as the main task, accounting for 85%, that is, the set of questions is mainly to examine the recognition ability, rather than understanding or higher-level cognition, which affects the effectiveness of the set of questions.

Secondly, the cognitive dimension examined in the reading part of the test questions is more inclined to recognition rather than understanding or higher-level cognition. Understanding dimensional questions accounts for about 47%, and recognition questions accounts for about 53%. The change of cognitive dimensionality does not show a gradient division, and the requirements for candidates are low, and the probability of correct answers is high, which affected the validity of the test questions.

In addition, the cloze part is not difficult to finish. The cloze part is to test the dimensions of identification, so it is not hard to set the questions, that is, it is not conducive to judging the students with good English proficiency.

Finally, the writing part covers 10 skills related to the rating scale. Among them, the four-level ability accounts for 60%, the three-level ability accounts for 10%, the five-level ability accounts for 20% and the second-level ability accounts for 10%. The results of sub-table comparison are contrary to the general table, and the four-level ability survey accounts for 60%. Spelling mainly examines students' mastery of key words. Translation examines students' application of key phrases.

3.4.3. Writing Part Improvement Measures

After the systematic analysis of each part of this set of questions combined with the scale, this study takes the writing part as an example and modifies it as a sample. Sample questions discuss traditional festivals in the form of web pages, so as to meet the description of Level 4 of the general table of written expression ability in the scale. Sample questions are as follows:

The English page of China Daily asks middle school students to introduce the Dragon Boat Festival. The main points are as follows:

1. Do you know why we celebrate the Dragon Boat Festival?
2. How do you think we should carry forward the culture of Dragon Boat Festival?

Reference vocabulary: Dragon Boat Festival; The lunar calendar

Notes:

1. The number of words is about 100;
2. Please answer the above questions and fill them in the corresponding website sections, from which we will select the most representative website pages and publicize them.

4. Conclusion

Based on the standard, the content of English academic achievement test of grade one in a middle school is marked. Through comparative analysis, the shortcomings of the test are obtained. In this study, the writing expression was adapted to make it more in line with the description of China's Standards of English Language Ability. This provides a reference for the primary and secondary school writing expression to match the scale better. Follow-up research can further improve other test by combining the scale, verify the construct validity of test by using item reflection theory and make corresponding improvements.

Acknowledgments

This paper is part of the research results of British Council Assessment Project: English Language Ability from the Perspective of Curriculum Ideology" supported by NEEA & BC English Assessment Research Grants (2020), — Examination Centre of the Ministry of Education. Project leader: Kangzhou Peng

References

- [1] Liu Jianda, 2015a, Some Thoughts on Developing Framework for English Language Proficiency[J], China Examinations (1): 7-11.
- [2] Lin Huiqing, 2016, Improve Citizen's Language Ability through Building a National Testing System [J], China Examinations (12); 3-4.
- [3] Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Strasbourg, France: Council of Europe (Modern Languages Division). Retrieved 10 November 2018 from <https://www.coe.int/en/web/common-european-framework-reference-languages>
- [4] Lowiea, W.M, Hainesa, K.B.J. & Jansmaa, P.N. (2010). Embedding the CEFR in the academic domain: Assessment of language tasks. *Procedia Social and Behavioral Sciences*, 3, 152-161 <https://doi.org/10.1016/j.sbspro.2010.07.027>
- [5] Swan, M. (2014). English Profile Studies 1. Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework. English Profile Studies 2. Language Functions Revisited: Theoretical and Empirical Bases for Language Construct Definition across the Ability Range. *English Language Teaching Journal*, 68 (1), 89-96. <https://doi.org/10.1093/elt/cct075>
- [6] Harsch, C. (2017). Proficiency. *English Language Teaching Journal*, 71(2), 250-253. <https://doi.org/10.1093/elt/ccw067>
- [7] Byram, M., & Parmenter, L. (Eds.) (2012). *The Common European Framework of Reference: The globalisation of language education policy*. Bristol, UK: Multilingual Matters.
- [8] Wu, J. (2012). Policy perspectives from Taiwan. In M. Byram & L. Parmenter (Eds.) (2012), *The Common European Framework of Reference: The globalisation of language education policy* (pp. 213-223). Bristol, UK: Multilingual Matters.
- [9] LIU Jianda. Validating China's Standards of English Language Ability[J]. *Modern Foreign Languages*, 2021, 44(01): 86-100.
- [10] HE Lianzhen, CHEN Dajian. Investigating the inner structure of China's Standards of English: Descriptive scheme and salient features of listening ability descriptors[J]. *Foreign Language World*, 2017(04): 12-19.
- [11] Mu Lei, Zhang Rong, Chen Guuangjiao. The correlation between listening competence and interpreting competence of BTI: An empirical study based on CSE self-assessment scales[J]. *Foreign Language Education*, 2021, 42(03): 88-93.
- [12] WU Zunmin. 2019. Standard-based Achievements: Linking China's Standards of English to content Standards[J]. *China Examinations*, (5): 36-44.
- [13] Wang Xianyi. (2008). *Study on Curriculum Standards-based Achievement Assessment Procedures*. East China Normal University.
- [14] Chen Ruisheng. (2009). New Thoughts on Academic Evaluation Based on Curriculum Standards. *Jiangxi Education* (35), 8-9. doi:CNKI:SUN:IWAS.0.2009-35-006.
- [15] Cui Yunhuo, Xia Xuemei. (2007). On the Evaluation of Students' Academic Achievement Based on Curriculum Standards. *Curriculum, Teaching Material and Method*(01), 13-18.
- [16] Wang Lei. (2007). Practice and Enlightenment of PISA2006 China Test Research. *Contemporary Educational Science*(Z1), 80-82.

Appendix

Part 1 Listening

	Overall scale for listening comprehension(CSE 2,CSE3,CSE 4)	Actual test points	Cognitive dimension	Frequency
China's Standards of English Language Ability(Overall Scales)	Can understand spoken language delivered at a normal speed on general topics that are of personal interest (e.g. speeches, news reports, talks); distinguish primary from secondary information based on discourse features; and grasp the main idea.(CSE 4)	Listen to long conversations and understand why the incident happened(8)	Understand	1
		Listen to the monologue and extract the essence of the text(17)	Identify	1
	Can understand short speech (e.g. talks, discussions, announcements) delivered with standard pronunciation at a slow but natural speed; and obtain key information with the help of stress, intonation, background knowledge, and contextual information.(CSE3)	Listen to monologues and understand the relationship between two of them from the information of multiple characters(19); Listen to short conversations, and guess the price of items (4), quantity information (9), and talk location (5); Listen to long conversations, and understand the speaker's views on related characters (12), and guess weight information (13)	Understand	6
		Listen to short conversations and use voice and intonation to obtain key information, such as numbers (1), appearance (16), character hobbies (3) and the reason for the event (2); Listen to long conversations, and extract detailed information, such as food (11), event information (14, 15), and extract the person in the event (7), the time of the event (6) and reason (10); Listen to the monologue, extract the way the event occurred (18), and extract the person's opinion (20)	Identify	12

(Subscales)	Can understand simple stories containing few low-frequency words when delivered at a slow speed and infer causal relationships among events(Understanding oral narration, CSE 2)	Listen to long conversations and understand why the incident happened(8)	Understand	1
	Can understand narratives about daily life and personal information when delivered slowly and grasp main ideas.(Understanding oral narration, CSE 2)	Listen to the monologue and extract the essence of the text(17)	Identify	1
	Can obtain key information from speeches or talks articulated clearly and delivered with standard pronunciation at a slow but natural speed Can understand short argumentation on familiar topics that is delivered in simple language at a slow but natural speed; and grasp the main idea.(Understanding oral argumentation, CSE 3)	Listen to monologues and understand the relationship between two of them from the information of multiple characters(19); Listen to short conversations, and guess the price of items (4), quantity information (9), and talk location (5); Listen to long conversations, and understand the speaker's views on related characters (12), and guess weight information (13)	Understand	6
		Listen to short conversations and use voice and intonation to obtain key information, such as numbers (1), appearance (16), character hobbies (3) and the reason for the event (2); Listen to long conversations, and extract detailed information, such as food (11), event information (14, 15), and extract the person in the event (7), the time of the event (6) and reason (10); Listen to the monologue, extract the way the event occurred (18), and extract the person's opinion (20)	Identify	12

Part 2: Reading comprehension

Standard		Detailed information	Cognitive level	Frequency
China's Standards of English Language Ability (Overall)	Can locate detailed information and summarize the main idea whilst reading different kinds of linguistically simple materials.(CSE4)	Read scientific articles and add title for it(35);Read stories and summarize the main idea(27);	Understand	2
		Read stories and identify wishes,(26),events(28,29); Read scientific articles and identify the reasons of events(30),way(32)	Identify	5
	Can understand the relationship between ideas by analysing the structures of sentences and discourse whilst reading materials of medium linguistic difficulty. (CSE4)	Read scientific articles and infer the implied meaning(34),understand the relationship of meaning(33); Read stories and understand the relationship of meaning(24)	Understand	3
	Can locate key information in linguistically simple practical forms of writing(e.g. letters, notices, signs)(CSE3)	Read practical writing and identify key information, such as things(22),route of the journey(21,23)	Identify	3
	Can understand the implicit meaning and summarize the main points of short, linguistically simple materials on familiar topics.(CSE3)	Read stories and understand the meaning of specific sentences(31).	Understand	1
	Can understand the relationship between points of information with the help of connectors in linguistically simple argumentative texts on familiar topics. (CSE3)	Read stories, locate key words and understand target information(25);	Understand	1
		Read exposition and identify cohesive devices of the discourse(36, 37, 38, 39, 40).	Identify	2
(Subscales)	Understanding written narration	Can recognise details (e.g. time, character, and place) in articles on social life, such as travel notes, written in relatively complex language.(CSE4)	Understand	3
		Can infer the author's intent from a narrative social life written in simple language.(CSE4)		
		Read narration, extract main events and the time.(28, 29, 30)		
		Read narration, understand the author's attitude.(31)		

	Understanding written exposition	Can understand the main points made in short popular articles.(CSE4)	Read exposition, extract key information.(32,33,34,35)	Identify	4
	Understanding written description	Can extract the main information about a scenic spot from a description written in simple language.(CSE3)	Read an introduction of the project, extract route, tools and time needed for journey.(21,22,23)	Identify	3
	Understanding written narration	Can extract the main information about a scenic spot from a description written in simple language.(CSE3)	Read story, understand positive information and values it conveyed.(24,25,26,27)	Understand	4

Part 3: Language

China's Standards of English Language Ability(Overall Scales)	Overall organizational competence	Can use appropriate vocabulary to describe events and define concepts(CSE 4)	Read the story and identify the description of the specific event(45, 46, 51, 55)	Identify	4
Subscales	Cohesion competence	Can use words to express contrast, addition, and/or other logical relationships(CSE 4)	Read stories and identify concessions (49)	Identify	1
	Vocabulary competence	Can understand basic vocabulary for daily topics relating to study, work, travel, and current affairs(CSE 4)	Read stories, and identify topics of physical condition (44, 50), emotional topics (44, 50), and emotional topics (41, 42, 43, 46, 47)	Identify	9
		Can use common adverbs to express time, degree, location, direction, sequence, and frequency (CSE 3)	Read stories, and identify frequency, degree, time, and adverbs(50, 54)	Identify	2

Part 4 Writing Spelling words

China's Standards of English Language Ability(Subscales)	Vocabulary Competence	Can command the meaning of core vocabulary, its common usage, and its meaning in specific contexts.(CSE 4)	66, 67, 68, 69, 70, 71, 72, 73, 74, 75	Identify
		Can understand basic vocabulary for daily topics relating to study, work, travel, and current affairs.(CSE 4)	66, 67, 68, 69, 70, 71, 72, 73, 74, 75	Identify
		Can flexibly use vocabulary to describe objects, behaviours, and features, as well as to define concepts.(CSE 4)	66, 67, 68, 69, 70, 71, 72, 73, 74, 75	Identify

Translation

China's Standards of English Language Ability(Overall Scales)	Overall written expression	Can describe in simple terms the main features of people or familiar objects in response to prompt(s) (e.g. words and examples). (CSE 4)	76, 77, 78, 79, 80	Identify
China's Standards of English Language Ability(Subscales)	Writing expression strategies	Can describe persons, objects, and places with the aid of prompts or from his/her own imagination(CSE 2)	76, 77, 78, 79, 80	Identify

Written Expression

China's Standards of English Language Ability(Overall Scales)	Overall written expression	Can briefly discuss familiar social and cultural matters (e.g. traditional festivals and customs) through social media (e.g. email and webpages).(CSE 4)	Write a letter to introduce the Dragon Boat Festival in China (time, sense, customs, etc.)	Identify
		Can express opinions on topic(s) he/she is familiar with, using some evidence to support his/her viewpoint(s) in a relatively persuasive manner(CSE 4)	1. Thanks to Rick for his hospitality 2. Welcome him to China	Identify
China's Standards of English Language Ability(Subscales)	Written description	Can describe a familiar scene or setting (e.g. a traditional festival or a celebration) in a detailed manner.(CSE 5)	Write a letter to introduce the Dragon Boat Festival in China (time, sense, customs, etc.)	Identify
		Can briefly describe personal activities (CSE 4)	Suppose you are Li Hua. You have just	Identify

			returned from a five-day tour in Britain. Thank your friend Rick for his hospitality.	
	Written interaction	Can briefly outline Chinese culture (e.g. traditional festivals and customs).(CSE 4)	Write a letter to introduce the Dragon Boat Festival in China (time, sense, customs, etc.)	Identify
		Can write letters to familiar people to share mutual experiences(CSE 3)	Suppose you are Li Hua. You have just returned from a five-day tour in Britain. Thank your friend Rick for his hospitality. Knowing that he is very interested in Chinese culture, especially the upcoming Dragon Boat Festival, please write a letter to introduce Chinese Dragon Boat Festival.	Identify