

Study on Content Validity of English Cloze in Guangdong Senior High School Entrance Examination

Author's Names: Ziqiao Su

Affiliations: Nanshanhu Experimental Middle School, Nanhai District, Foshan City, Guangdong Province, China

Postcode: 528000

Email: suziqiao99@163.com

Phone number: +86 18613099430

Corresponding Author: Ziqiao Su

Email: suziqiao99@163.com

Phone number: +86 18613099430

ABSTRACT

The study explored the content validity of English cloze based on five sets of English cloze tests from 2014 to 2018. The author revealed that the content validity did not meet curriculum standards, except for vocabulary, and the validity of grammar, themes and functions was low, with the knowledge of grammar being incompletely covered. The knowledge of grammar, range of themes and functions should be expanded in future cloze to better evaluate students's language ability

Keywords: cloze test, content validity, grammar, themes, functions

1.0 INTRODUCTION

Cloze can be divided into various types according to different classification criteria. Among these types, selective cloze is the most common. An example of selective cloze is English cloze, used in Guangdong senior high school entrance examinations. As a comprehensive test with strong objectivity, cloze is widely used in all kinds of English tests and can examine the language knowledge of students as well as their discourse comprehension (Li, 1997). Cloze tests are given to evaluate the ability of students to use their language knowledge comprehensively. Guangdong senior high school entrance examination test paper should not only accordance with the syllabus, but also the curriculum standard. However, the syllabus for candidates of junior high school academic level has recently been cancelled. In 2019, Ministry of Education of the People's Republic

of China (MOE) explained that teaching should be conducted according to the curriculum standards. This means that teachers should pay more attention to the content validity of tests in accordance with the curriculum standard.

Validity has always been an important criterion in evaluating English language proficiency tests, and as such, it is important to discuss what validity means (Abedi et al., 2012). Researchers have defined the word in different ways. Henning (1987) defines validity as to how to measure the test paper properly. Messick (1989) described validity as a reasonable explanation of test scores based on evidence and theoretical basis. According to Hughes (2003), a test is said to be valid if it accurately measures what it intends to. Validity refers to the degree to which the measured result reflects the content to be investigated. Content validity is one kind of validity. In 2000, Alderson, Clapham, and Wall believed that content validity was the most important part of internal validity, as the higher the content validity, the more accurate the content to be tested was. Content validity is an important criterion for evaluating the quality of a test paper. A common way to ensure the content validity of a test is to analyse the content and compare it to a standard test, be it a test specification, formal syllabus, or curriculum standard (Alderson et al., 2000).

After searching online for articles related to English cloze tests and content validity, I have found that the current research on content validity mainly involves studying English reading, while there is very little research on the validity of English cloze. In addition, most people use the examination syllabus for analysis, rather than on the basis of curriculum standards. Thus, it is necessary for us to study whether the cloze tests meet the requirements. The overall objective of the compulsory education English curriculum consists of five parts: language skills, language knowledge, emotional attitude, learning strategies, and cultural awareness. Of these, language knowledge, in particular is an important part of language ability. English cloze test papers will be compared with the English curriculum standards of compulsory education to test whether the questions within them meet the English curriculum standards for compulsory education (2011), as well as to analyse the content validity of cloze tests in Guangdong senior high school entrance examinations. Thus, this essay is mainly focused on how the content validity of cloze tests matches the curriculum requirements in terms of language knowledge.

The common method for checking research content validity is to analyse the test content and compare it with the test standard. Bachman and Palmer's framework can be used to describe existing tests and to provide testers with a basis for adjusting test content. Hughes (2003) claims that 'In order to judge whether or not a test has content validity, we need a specification of the skills or structures, etc. that it is meant to cover' (p. 26). Test specifications have to be translated into test items, and an important part of the judgement process in testing is deciding whether the test adequately realizes the specifications (Alderson, 2011). Based on this, the researchers agree that a detailed specification is needed to describe the scope of the test. Bachman and Palmer's useful framework helped practitioners to move beyond thinking of validity as a characteristic of a test to a

more contextualized (Abedi et al., 2012). The framework is presented in Table 1.

This essay is divided into four parts. In the first part, the cloze , relevant research on validity and content validity, and Bachman and Palmer’s framework are introduced, followed by a discussion on the necessity of research. The second part displays the method the study has used. The third part presents the findings and discussion of the study, from which the last part draws a conclusion.

Table 1. Bachman and Palmer’s Framework

Characteristics of the setting	
<i>Physical characteristics</i>	
<i>Participants</i>	
<i>Time of task</i>	
Characteristics of the test rubrics	
<i>Instructions:</i>	Language (native, target) Channel (aural, visual) Specification of procedures & tasks
<i>Structure:</i>	Number of parts/tasks Salience of parts/tasks Sequence of parts/tasks Relative importance of parts/tasks Number of tasks/items per part
<i>Time allotment</i>	
<i>Scoring method:</i>	Criteria for correctness Procedures for scoring the response Explicitness of criteria and procedures
Characteristics of the input	
<i>Format:</i>	Channel (aural, visual) Form (language, non-language, both) Language (native, target, both) Length Type (item, prompt) Degree of speededness Vehicle (‘live’, ‘reproduced’, both)
<i>Language of input:</i>	Language characteristics Organizational characteristics

Grammatical (vocabulary, syntax, phonology, graphology)
 Textual (cohesion, rhetorical/conversational organization)
 Pragmatic characteristics
 Functional (ideational, manipulative, heuristic, imaginative)
 Sociolinguistic (dialect/variety, register, naturalness,
 cultural references, & figurative language)
 Topical characteristics

Characteristics of the expected response

Format: Channel (aural, visual)

Form (language, non-language, both)

Language (native, target, both)

Length

Type (selected, limited production, extended production)

Degree of speededness

Language of expected response

Language characteristics

Organizational characteristics

Grammatical (vocabulary, syntax, phonology, graphology)

Textual (cohesion, rhetorical/conversational organization)

Pragmatic characteristics

Functional (ideational, manipulative, heuristic, imaginative)

Sociolinguistic (dialect/variety, register, naturalness,
 cultural references & figurative language)

Topical characteristics

Relationship between input and response

Reactivity (reciprocal, non-reciprocal, adaptive)

Scope of relationship (broad, narrow)

Directness of relationship (direct, indirect)

(Bachman & Palmer, 1996, pp. 47–57)

2.0 METHOD

2.1 Materials

Five sets of English cloze tests for the high school entrance examination of Guangdong province from 2014 to 2018 were selected. There were five small articles, with ten blanks in each article and four choices associated with each blank. Only one of each four choices would be a correct answer. There was a total of 50 cloze questions, all of which were single choice questions.

2.2 A New Framework for Assessing the Content Validity of Cloze Tests

The framework of Bachman and Palmer's framework was adjusted by combining it with the curriculum standards of compulsory education. Bachman and Palmer (1996) said that the order of the framework will not necessarily be consistent with the target, and thus needs to be adjusted appropriately.

In the course standards of compulsory education, there are nine levels in total, with the fifth level being the goal standard for ninth-grade students to reach. As has been previously stated, language knowledge is an important component of pragmatic competence. Ninth-grade students in compulsory education stage aiming for level five should master basic knowledge of the English language, vocabulary, grammar, functions, and topics. Cloze tests do not measure knowledge of phonetics, so this aspect cannot be discussed. Vocabulary, grammar, and topic are part of the input section, while function is related to expected response. Setting and rubrics already fall under the requirements of the problem described earlier. Therefore, this framework mainly considers input and expected response, as seen in Table 2.

Table 2. A New Framework for Assessing the Content Validity of Cloze Tests

Characteristics of the input	Vocabulary Grammar Topic
Characteristics of the expected response	Functions

3.0 FINDINGS and DISCUSSION

3.1 Characteristics of the Input

3.11 Vocabulary

In vocabulary, curriculum standards require students to understand and comprehend the basic meanings of words as well as their meanings in specific contexts. In order to check whether vocabulary can be understood by students, test the readability of several articles if you have any at hand to determine the difficulty they may have in reading them (Liontas & DelliCarpini, 2018). Using the level of readability to measure the readability of words. If these words can be understood by the examinee, the validity is high. Flesch Reading Formula quantifies the legibility of the input

text in terms of its linguistic form by taking the word length and sentence length as the main factors. The higher the score, the easier an article is to read.

Table 3 shows a readability yardstick. As can be seen, the vocabulary is of moderate difficulty suitable for the reading level of ninth-grade students. The continuation of fixed-ratio cloze tests in educational testing may be illustrated by evidence showing the cloze procedure to be a reliable standard of readability (Gellert & Elbro, 2013). The Flesch Reading Ease score is directly between 70 and 90, indicating that these articles are easy or fairly easy to read. The Flesch Kincaid grade 5 is equivalent to the reading level of a child aged 8–9 in the United States, and grade 6 is equivalent to the reading level of a child in grade 6 in the United States. Alderson (2011) believes that foreign language readers need to achieve a certain level of reading ability in order to convert their first language into a second language.

Students are able to understand the context of the passage, and as such, meet the curriculum standards. Aside from the difficulty in 2016 and the simplicity in 2015, the rest of the readability is distributed at a relatively stable level. Chinese students can be less demanding than American students as they are not native speakers of English, and in this sense, cloze tests meeting the requirements of the curriculum standards which has high validity.

Table 3. Readability Yardstick

Score	Description of Style	Potential School
90–100	Very Easy	4 th grade
80–89	Easy	5 th grade
70–79	Fairly Easy	6 th grade
60–69	Standard	7 th or 8 th grade
50–59	Fairly Difficult	Some High School
30–49	Difficult	High school
0–29	Very Confusing	College

(Source: <https://readabilityformulas.com/flesch-reading-ease-readability-formula.php>)

Table 4. 2014–2018 Text Word Test Results

Year	2014	2015	2016	2017	2018
Flesch Reading Ease score	84.3	82.7	78.1	85.8	86.9
Flesch–Kincaid Grade Level	5	4.6	6.5	5.5	5.2

3.12 Grammar

The syllabus (2011) requires students to understand the items within the list of grammar items and be able to use them in a specific context. Students should not only master their knowledge of grammar but also know how to use it correctly in specific situations. To test whether students can use grammar in an actual situation, the test points should be evenly distributed and be consistent with the requirements of the curriculum standards. In order to maintain the accuracy of the test, the test points should be uniform, which indicates that the content validity of a question is high.

Cloze tests mainly investigate the use of notional words and function words. As can be seen in Table 5, the test points are not evenly distributed. In Table 5 there are nine test points in total, covering only four to six points on average. Only six test points are examined, including nouns, pronouns, verbs, conjunctions, adjectives, and adverbs. The average coverage rate was 57.82%. In 2014, there were at least four test centres covered, accounting for 44.5%, and in 2017 and 2018, there were six, accounting for 66.7%. Nouns, verbs, and adjectives are the most frequently examined, and there are tests concerning them every year. As seen in Table 6, three adjectives were examined in 2014, accounting for 30% of the questions, and two were examined from 2015 to 2018, accounting for 20%. In 2014, verbs were examined the most, with five used, accounting for 50%. In other years, three to four were used. Adjectives received the most attention in 2016 with three examined (accounting for 30%), and had between one and two used in all other years. Pronouns and adverbs are less frequently examined, with the former only being tested in 2017 and 2018. In terms of conjunctions, there was no examination in 2014 and only one examination in all other years. Adverbs were examined annually from 2014 to 2018. Other numerals, articles, prepositions, and prepositional phrases were not examined.

Grammar coverage points did not meet the requirements of the syllabus, did not see the examination of complex sentences, subject-verb agreement and other test points. The grammar of cloze does not meet the content validity. Siddiek (2010) found that when the exam lacked content validity, most of the teaching would be focused in the techniques of examination rather than doing real teaching. Therefore, when considering only the distribution of cloze test grammar points, the content validity does not appear to be high as the distribution is uneven.

Table 5. Examination Site Coverage Analysis Form

Year	2014	2015	2016	2017	2018
Noun	3	2	2	2	2
Pronoun	0	0	0	1	1
Numeral	0	0	0	0	0
Verb	5	4	3	3	4
Conjunction		1	1	1	1
Adjective	1	2	3	2	1
Adverb	1	1	1	1	1
Article	0	0	0	0	0
Prepositions and prepositional phrases	0	0		0	0
Percentage	44.50%	55.60%	55.60%	66.70%	66.70%
The average percentage					57.82%

Table 6. Inspection Proportions

Year	2014	2015	2016	2017	2018
Noun	30%	20%	20%	20%	20%
Verb	50%	40%	30%	30%	40%
Adjective	10%	20%	30%	20%	10%

3.13 Topic

There are a total of twenty-four topics in the curriculum standard, which can be summarized into five topics which students should be familiar with. These topics are personal and school life, emotions, customs, hobbies, and science and culture.

As can be seen from Table 7, these five passages only cover two topics which are not comprehensive enough to meet the requirements of the course standards: personal life, accounting for 40%, and emotions, accounting for 60%. Hobbies, customs and habits, and Science and culture are not covered. To explore the reasons for this, it can be found that these five cloze essays are narrative essays which do not involve other articles such as expository texts, which may make it much easier for students to choose their answer. The difficulty of cloze tests can be explained by the reminder features (Trace, Brown, Janssen, & Kozhevnikova, 2017).

Therefore, in the past five years, the range of other topics is not widely distributed. Once students grasp a topic, they will practice that topic again and again. This cannot test the real level of students. Therefore, cloze tests need to expand the topic range and improve in content validity.

Table 7. Topics of Cloze Tests

Year	2014	2015	2016	2017	2018	Percentage
Personal and school life	√		√	√		60%
Emotions		√			√	40%
Customs						0%
Hobbies						0%
Science and culture						0%

3.2 Characteristics of the Expected Response

3.2.1 Functions

Function is a term used to describe the illocutionary force of what is said. (Weir, 2005, P. 78). English cloze reflects some functions to be expected response. For example, through a narrative English cloze passage, expected students response personal feelings, attitude tendency, comparison, etc. Weir (2005) added that the article needs to consider the functional purpose of the text, leaving readers guessing about the author's intentions. A cloze test not only measures students' vocabulary, but also some functional items. The curriculum standard has 11 functional items. The function of the project was observed under the guidance of an English teacher who teaches junior high school English in an educational institution. The distribution of functions involved in completion filling was classified in Table 8. The teacher was willing to instruct and signed the informed consent form.

It can be seen that only five of the eleven functions are covered, with the topic of cloze not being fully covered. The first three functions that were covered are Logical relations, Attitudes, and Emotions. Logical relations involved 26 items, accounting for 52%. Attitudes involved 10 items, accounting for 20%, while Emotions was covered with 8 items, accounting for 16%. These 5 cloze test just expecting students to respond to these 3 kinds of expected responses. However, Curriculum requirements emphasis testing should be placed on examining students' comprehensive language application ability, that is, students are expected to respond to cover most of the function. There are few investigations in Social Communications and Space, and almost no studies in Time, Measure, Features, Existence, and Occupation. This may be because these areas have been investigated in other questions but are not the focus here. The cloze test should meet the functional requirements of the course standards for student grasp of the text, analysis of context and logical reasoning ability. The reading ability tests assess the text function and comprehension(Kleijn, Pander & Sanders, 2019). In the functional aspect, the investigation is not thorough enough.

In general, the coverage of the function is relatively small, which is not in accordance with the requirements of the curriculum standards for skilled use of 11 functions to read. Therefore, the content validity is not very high.

Table 8. Function and Distribution of Five Sets of Cloze Tests

	2014	2015	2016	2017	2018	Total (50)	Percentage
Social Communications	0	1	0	1	2	4	8%
Attitudes	6	0	0	3	1	10	20%
Emotions	2	3	1	0	2	8	16%
Time	0	0	0	0	0	0	0%
Space	0	0	0	0	1	1	2%
Measurement	0	0	0	0	0	0	0%
Features	0	0	0	0	0	0	0%
Existence	0	0	0	0	0	0	0%
Comparison	1	0	0	0	0	0	0%
Logical relations	1	6	9	6	4	26	52%
Occupation	0	0	0	0	0	0	0%

4.0 CONCLUSION

This essay has analysed how the content validity of cloze tests matches the curriculum requirements in terms of language knowledge. It concludes that the content validity of cloze tests and the curriculum requirements do not match very well, with only vocabulary meeting the curriculum requirements. However, there is low content validity in terms of grammar, topic, and function. The coverage of grammar knowledge points is uneven. Therefore, grammar should be included in the investigation such as complex sentences and subject-predicate agreement. The coverage of topics does not meet the requirements of the course standards and should thus be expanded, as should the coverage of the investigation function.

However, the sample size is small as only five cloze tests have been chosen. This cannot completely represent the characteristics of Guangdong high school entrance examinations, making the findings subjective to an extent. Besides, not enough teaching experiences, break away from teaching practice, simply analyse the passage take incomprehensive consideration.

References

- Abedi, G., Antón, M., Brown, A., Brown, J. D., Chapelle, C. A., Chapelle, A. D. ... Davies, A. (2012). *The Routledge handbook of language testing*. Fulcher, G. & Davidson, F. (Ed.) New York: Routledge
- Alderson, J. C., Clapham, C., & Wall, D. (2000). *Language Testing Construction and Evaluation*. Beijing: Foreign Language Teaching and Research Press.
- Alderson, J. (2011). *A lifetime of language testing*. Shanghai: Shanghai Foreign Language Education Press
- Bachman, L. F., & Palmer A, S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16–28. doi:10.1177/0734282912451971
- Henning, G. (1987). *A guide to language testing: Development, evaluation and research*. Beijing: Foreign Language Teaching and Research Press.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed). Cambridge: Cambridge University Press.
- Kleijn, S., Pander Maat, H. L. W., & Sanders, T. J. M. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing*, 36(4), 553–572. doi:10.1177/0265532219840382
- Li, X. (1997). *The science and art of language testing*. Changsha: Hunan Education Press.
- Liontas, John I., International Association, TESOL, & DelliCarpini, Margo. (2018). The TESOL encyclopedia of English language teaching || cloze tests.
- Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational Measurement*, 3rd edn. NY: New York.
- MOE : Ministry of Education of the People's Republic of China [Website]. (2019) http://www.moe.gov.cn/fbh/live/2019/51594/mtbd/201912/t20191202_410397.html

- Siddiek, A. G. (2010) The Impact of Test Content Validity on Language Teaching and Learning. *Asian Social Science* 6 (12)
- Trace, J., Brown, J. D., Janssen, G., & Kozhevnikova, L. (2017). Determining cloze item difficulty from item and passage characteristics across different learner backgrounds. *Language Testing*, 34(2), 151–174. doi:10.1177/0265532215623581
- Weir, C. J. (2005) *Language testing and validation: An evidence-based approach*. NY: New York