

Investigating the Content Validity of the Junior Middle School Entrance English Testing

Rong Xie¹

¹ Department of Foreign Studies and Tourism, Tongren Preschool Education College, Guizhou, China

Correspondence: Rong Xie, Department of Foreign Studies and Tourism, Tongren Preschool Education College, Tongren, Guizhou 554300, China.

Received: February 5, 2018

Accepted: March 1, 2018

Online Published: March 8, 2018

doi:10.20849/jed.v2i1.334

URL: <https://doi.org/10.20849/jed.v2i1.334>

The research is financed by Research on Humanities and social sciences of Guizhou Provincial Department of Education. Project No: 2016ZC137. Young project. The title of the project: On Games in Teaching English to Young Learners.

Abstract

The primary purpose of this study was to investigate the content validity of the Junior Middle School Entrance English Testing. To understand the extent to which the test is consistent with the National Curriculum Standards, the qualitative and quantitative methods were implemented in this study, through the comparative analysis of the whole article's items, and combined with the characteristic of the communicative language task to seek out whether the test content is relevant to the test objective, and to what extent it is representative. The most important thing is to find out whether the test content is suitable for the test object and to what extent. The results showed that the Junior Middle School Entrance English Test to some extent matches the syllabus, meanwhile, there are still some issues to exist. Finally, some suggestions were put forward by the author.

Keywords: test, content validity, young learners

1. Introduction

1.1 Background

With the globalization of economy and tourism, English is acknowledged as the "world's language". Young learners also have emerged an upsurge of learning English. In some case, how to evaluate children's English proficiency effectively has been identified as a hot topic in the field of contemporary education. The current research mainly concentrates in what way to evaluate young learners' English proficiency. Different researchers have different arguments relate to the ways of young learners' assessment. Some test experts thought that the children should be conducted to evaluate via self-assessment, portfolio, formative assessment and task-based assessment. But some other researchers argued that based on the standardized test the norm-referenced test are regarded as a good way to pupils. Douglas (2012) argued that numerous schools have utilized standardized test in order to assess learners' ability and progress in primary school.

Bachman (1999) defined that a standardized test is designed in such a way that the questions, conditions for administering, and scoring procedures refer to a standard manner. As an assessment method, the standardized test plays an important role in English as a foreign language curriculum evaluation as well as learners' evaluation. In my local context, the Junior Middle School Entrance English Test is as a standardized test. The type of the test belongs to the paper-and-pencil test. The aims are to assess the pupils' English whether or not reach the level of junior middle school. McKay (2006) emphasized that the test for young learners are different from adults. Although considerable research has been devoted to the content validity of the standardized test, rather less attention has been contributed to the content validity of the young learner English test. The test should be considered in young learners' continuous learning potential ability and communicative language ability. It is only as a reference to diagnose whether the learners possess the competence of learning English. In short, the test aims to probe learners whether or not have genius ability to learn a foreign language in the junior middle school. Additionally, it's also as an important indicator to students whether can successfully enter into junior middle school with an excellent reputation. Namely, the test is designed based on the requirements of the National

Primary School English Teaching Syllabus. The tests are administered by departments of local education, through comparison with the requirement of the syllabus, through the comparative analysis of the whole article's items, and combined with the characteristic of the communicative language task to seek out whether the test content is relevant to the test objective, and to what extent it is representative. The most important thing is to find out whether the test content is suitable for the test object and to what extent, etc.

1.2 The Purpose of the Study

Bachman (1990) put forward that a good paper-pencil test means the test should be tested in a sense. Macky (2006) argued that a good paper-pencil test is considered that the result has been a strong positive backwash to stakeholders. In contrary, a bad paper-pencil test means to the extent which has not tested the purpose what the tester wants to. It is regarded that the test has seriously negative backwash to them. The validity of a test was regarded as one of the most significant indicator of the measure for language testing (Lynn, 1986). In the past decade, many researchers of foreign language test content validity's research focused on the Senior Entrance Test, the College Entrance Test, CET4, etc. However, the content validity of the Junior Middle School Entrance English Test is still a vacancy, whether the validity of the test fits for learners' practical condition, or whether the contents of the test match the syllabus. In short, the primary purpose of this study was to investigate the content validity quality of the Junior Middle School Entrance English Testing.

1.3 Significance of the Study

In order to accurately assess English as foreign language development of primary graduated learners during the period of elementary school. The Department of Education and school administrators developed a pilot test of one standardized assessment designed for this aim. The main significance of this article is to investigate the quality of the junior middle school entrance English test and offer empirical evidence related to its use as a standardized test. Investigating the content validity of the test whether or not can make an accurate judgement for young learners' English communicative ability. Bachman (1996) proved that the content validity is crucial to judge the test whether matches the requirements of the syllabus. Hence, this study mainly analyses the items of a set of paper-pencil tests of the Junior Middle School Entrance English Test (2016), which combined with a model of components of language *competence* (Bachman, 1990, p. 87).

1.4 The Structure of This Study

This study consists of six sections. Firstly, it is the introduction refers to my problematic issue. Secondly, some previous research about validity will be involved in this part. Additionally, the literature review will be argued through the types of validity, content validity and the communicative language ability. The author also needs to describe the test syllabus of the Ministry of Education. Thirdly, it is the methodology relates to the materials of the test. Fourthly, the author will use the framework of communicative task characteristics as a reference, then to continue analyzing the listening part, the multiple choice characteristic, writing characteristic of the whole test paper. Fifthly, the findings and recommendations will be involved. Sixthly, this study will be summarized.

2. Literature Review

2.1 The Concept of Validity

Validity is the central concept in testing and assessment. Henning (1987) defines validity as follows: A test is regarded as validation to the extent that evaluates what it is supposed to measure. He claims that one of the potential problems in the test is the misuse of the test. Namely, administrator uses a test for an intention for which it was not intended. The validity of the test is identified invalidity after the validity of use for that purpose needs to be demonstrated and established. Hughes (2003, p. 22) confirmed that "Validity" in testing and assessment has traditionally been understood to mean discovering whether a test "measure accurately what it is intended to measure". However, Gronlund (1998) argued that the validity refers to the evaluation's purpose, whether the measurement of the test's results are appropriate, effective and significant. Researchers in different fields have different definitions about the "validity". Wagner and Braun (1998) put forward the validity in quantitative research as "construct validity". Construction is the initial notion, the concept, hypothesis, or the question that confirmed which data is to be collected and how it is to be gathered.

2.2 The Types of Validity

In the early stages of the validity study, most testers have identified three main types of validity. Namely, construct validity, empirical validity and rational validity. Hughes (2003) argued that validity should be divided internal validity and external validity, with the distinction being that internal validity refers to studies of the perceived content of the test, while external validity relates to studies comparing learners' test scores with their own abilities. Alderson and Wall (1995) also put forward that internal validity includes face validation, response

validation and content validation. Henning (1987, p. 96) indicated that the response validation is through part of test takers' self-report or self-observation are used to interpret how they respond to assessment items. External validity is also called criterion validity. The most common types of external validity are predictive and concurrent validity. In a word, different test experts have different ways of classifying "validity".

Through the analysis of the previous study to what extent the validity should be divided into four "types", each type of validity was referred to a kind of evidence that a test was valid. The first one is face validity which means that the surface appears to be valid, whether the text looks like it measures what it is supposed to measure. Wainer and Braun (1988) also argued that the face validity to what extent non-testers, such as learners and administrators to evaluate the value of the test. The second one is criterion-oriented validity which concludes predictive validity and concurrent validity. Predictive validity is the term used when the test scores are used to predict some future criterion, while if the scores are used to predict a criterion at the same time the test is given, it was called concurrent validity. The third one is construct validity which based upon an investigation of the psychological constructs or characteristics of the test. Construct validation focuses on what the test scores actually mean. Alderson and Wall (1995) confirmed that construct validity is a form of assessment validation to what extent it is successfully based on test potential theory. Finally, one of the most significant aspects is content validity which is also called rational validity. Content validity is based upon an analysis of the body of knowledge surveyed (Fulcher & Davidson, 2008).

2.3 Content Validity

Alderson and Wall (1995) pointed out that the so-called content validation is based on the tester, linguistic or expert's judgement of the test content. A test will be confirmed to have content validity if its content is composed of a representative sample of language structures, skills, etc. For instance, grammar test must be made up of items directly related to the knowledge or control of grammar. Henning (1987) argued that a test paper only has adequate representative sample of the items which do not ensure content validity. A valid content test paper needs to be included in a proper sample of the relevant structure. Hughes (2003) had the same viewpoint with the above. Of course, the relevant structures will depend upon the purpose of the test. Different English level learners need to be designed for different test contents. In short, stake-holder would not expect a set of paper-pencil test for intermediate learners to contain just the same set of structures as one for advanced learners or backward learners. If items are not a good sample, the results of testing will be misleading.

Additionally, Wainer and Braum (1988) argued that one of the most important issues needs to be considered by the test developer is coverage. Probably, it is not expected that everything in the syllabus and curricula will always appear in a set of test, because a single test can't contain everything in the teaching course. But it will provide the test constructor with the basic principles of Syllabus and Curricula. Therefore, Hughes (2003) indicated that another critical factor should be considered in the domain of content validity. What is the score weight of the test content or item? How to allocate knowledge points the Syllabus Required? In a word, what is the value of testing? Not only the author has to consider its representativeness and relevance, but also needs to measure its coverage and weightings in testing language (Hughes, 2003).

2.4 The Communicative Language Ability (CLA)

Bachman (1990, p87) described three components of CLA: Namely, language competence, strategic competence and psychophysiological mechanisms.

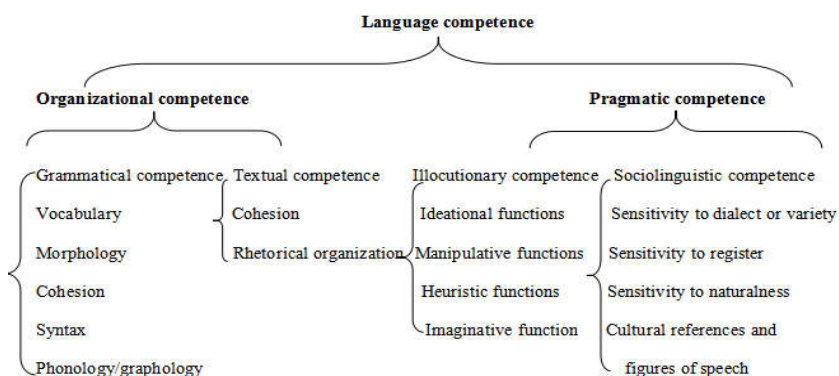


Figure 1 Components of language competence(Bachman, 1990, p87)

Grammatical competence is involved in knowledge of lexis, morphology, syntax, phonology and graphology, and knowledge of sentence formation. Bachman proved that textual competence means knowledge of the conventions for joining utterances together to come into being a text according to the rules of cohesion and rhetorical organization. Bachman (1998) pointed out that textual competence is also involved in conversational language usage, units of language larger than sentences. Cohesion is the way of explicitly marking sentence relations and conventions governing the ordering of old and new information in discourse. He also identified that rhetorical organization contains the whole conceptual structure of text and is referred to the effect of a text on the language user (narration, description, comparison, classification and process analysis). To this point, Fulcher and Davidson (2007) have illustrated the same list.

Pragmatic competence means knowledge about the relationships between utterances and the acts/functions by the speakers or writers intends to perform through them and about characteristics of context of language use.

Bachman and Palmer (2010) defined that illocutionary competence means functions of language. Sociolinguistic competence means sensitivity control of the language use that is demonstrated by the particular language use context.

2.5 The Test Syllabus

According to the requirement of the National Primary English Curriculum Standard, the test syllabus is based on the Curriculum Standard. In the stage of Compulsory Education, English language teaching and learning are divided into nine proficiency-based levels in China. English Curriculum at the Primary Period was designed into two levels. Level 1 covers Grade 3 and Grade 4, whilst level 2 covers Grade 5 and Grade 6. The framework of the objective of the National English Curriculum is based upon five aspects. Namely, language knowledge, language skills, learning strategies, affect attitude, and cultural awareness. Here objectives for the level 2 will be emphasized at the following points:

- 1) To cultivate pupils' interests, self-confidence and positive attitude towards learning the second language.
- 2) To help children form good habits in learning a foreign language.
- 3) To foster the pupils' elementary ability to use English in daily communication context.
- 4) To lay a good foundation for spoken English and prepare for further study. The overall goal is to cultivate the young learners' language competence (the Department of Education, 2011).

3. Methodology

3.1 Participants

The participants of this study were 6th grade-level learners from a public primary school in Guizhou. The local Education Department provides students with a unified course-book. The test was administered at the end of the academic year, in July 2016.

3.2 Measure

The qualitative and quantitative methods were involved in this study. Jack (1979, p. 601) argued that more than one methods should be used in the study of the same phenomenon to obtain the feedbacks from the multiple angles. The author elicited the findings through the analysis of the qualitative document and combined with the quantitative data statistics. The aim of the junior middle school entrance English test is to assess the communicative language ability. The test was designed on account of the five components of the Syllabus. The content validity of this test paper will be analyzed via the framework of communicative task characteristics (Bachman & Palmer, 2010). In addition, to compare with the requirement of the syllabus, four issues as follows will be investigated by the author:

- 1) Is the content related to the test aim of the Junior Middle School Entrance English Testing? If yes, to which aspect?
- 2) Does the content of the test have representativeness?
- 3) Is the content of the test suitable for the test object? If yes, to what extent?
- 4) What is the score weight of the test contents in the whole set of paper?

3.3 The Framework of Communicative Task Characteristics

The framework is designed by Bachman and Palmer (2010) to measure the content validity of test paper more effectively.

The framework of communicative task characteristics	
1 Setting	Physical characteristic, participants, time of task
2 Rubric	Instructions: Language, channel, specification of procedures, procedures for producing assessment record, Structure: Number of part/tasks, salience of parts/tasks, sequence of tasks, relative importance of tasks, number of tasks, Time allotment/recording method: Type of record, criteria for correctness, explicitness of criteria and procedures for recording the response, recorders, procedures for recording the response
3 Input	Format: Channel, form, language, length, vehicle, degree of speededness, Language characteristics: Grammatical, textual, function, dialect / varieties, genre, register, idiomatic expressions, cultural reference, figures of speech, topical characteristics
4 Expected response	Format, language characteristics, topical characteristics
5 Relationship	Type of external interactivenss, scope, directness

Figure 2 The framework of communicative task characteristics (Bachman & Palmer, 2010)

3.4 The Framework of the Junior Middle School Entrance English Test

According to the primary English curriculum and combined with the framework of communicative task characteristics (Bachman & Palmer, 2010), a new framework of the Junior Middle School Entrance English Test was designed by the author.

The framework of the junior middle school entrance English test	
1 Setting	Physical characteristics: based on the full-time compulsory education English curriculum standard, Participants: young learners(age:11-12)and invigilators/administrators, Time of task: at the end of the academic year, in July 2016
2 Rubric	Test instruction: target language and test-taker's native language, Channel: written, aural, visual and oral, Structure: number of the parts/tasks ,salience of parts/tasks ,sequence of tasks, relative importance of tasks, number of task, Time allotment:50 minutes, Scoring method: score/objective scoring key
3 Input	Length: simple sentence, Language characteristics :simple grammar ,short text, ideational/imaginative function, Topic/genre: variable
4 Expected response	Types of response: written, English, selected response

Figure 3

3.5 The Item Types for the Junior Middle School Entrance English Test

A set of test paper included two major sections: Listening (40points) and writing (60points). The total score is 100 points. Test time is 50 minutes, listening including 4 sub-sections, 20 items. Writing has 6 sub-sections, 25 items, each test item worth 2 points, and composition section worth 10 points. Multiple-choices have 30 items, which accounted for the total score of 60%.

4. Analysis and Discussion

4.1 The Analysis of the Listening Parts

As we know that it's extremely important to improve young learners' listening ability in English teaching, so assessing learners' listening ability are also concerned by the test paper designer in terms of the Curriculum Standard. According to this test paper, we can see that listening part is accounted for 40% of the total score, to some extent which has emphasized the weightings of listening. For this section, the author would like to make an analysis from three aspects, namely, length, topic/genre, and authenticity of the materials.

4.1.1 The Analysis of the Length

“Length” can refer to the single words, phrases, sentences, paragraphs and discourse (Bachman & Palmer, 2010). The length of listening materials directly affects the test results. There is something to be sure of testing young learners to the extent which differs from middle school students. In comparison with adults, young learners have a shorter attention span. Based on this argument, the majority of the listening materials in the primary phase are composed of single words and very simple sentences. By calculating, the longest word in this test paper consists of maximum 6 letters, and the longest sentence in testing paper was composed of only 10 words. Therefore, the author doesn't have to use the software to analyze the length of the listening materials, via concrete the analysis of the results can be judged by the “length” in listening to some extent which has no impact on young learners.

4.1.2 The Analysis of the Topic and Genre

The Curriculum Standard (2011) shows that the language goals for two-level Grades in the listening part are that:

- 1) Be able to listen to simple words and sentences, to understand elementary tape recording materials with the help of the pictures and gestures.
- 2) Be able to listen to and understand a simple story with pictures.
- 3) Be able to understand simple questions during the classroom activities.
- 4) Be able to understand common instructions and requirements in daily life. Show his/her appropriate reactions to what teacher says.

The topic/genre in the listening part ^a									
Topic	greeting	hobby	request	shop	weather	introduction	thanks	people	thing ^a
Item 1		2	1			2			
Item 2		1		1				2	1 ^a
Item 3		3	1		1 ^a				
Item 4	1		4 ^a						
Total	1	6	6	1	1	2		2	1 ^a

Figure 4^a

From the Figure 4 the topic of the items nearly covered all related to the young learner's daily life within the list of syllabus. Especially the topic of “hobby and request” is more welcomed by administer. Because they can bring test taker into a real-word in order to inspire his/her interests to learn English. Bachman and Palmer (2010) claim that authenticity is critical to “demonstrate how performance on a given language test based on language use in a specific context rather than the language test itself”.

4.2 The Analysis of the Multiple Choice Characteristics

Multiple choice can work well for testing lower level skills (Vicky, Meg, & Jake, 2009). As the new standards are designed to develop the learners' overall quality, which pay more attention to the pupils' overall development in language skills, language knowledge, affects, learning strategies and culture awareness so as to lay a good basis for future study. Encourage the use of both formative and summative assessment to measure the children's English level. So language knowledge and language skills will be checked in the concrete content. In addition, probably the most distinct advantage of multiple choice is that scoring can be absolutely reliable, rapid and economical (Hughes, 2009, p.76). Hence, in some case, it is favored by the test developer. Meanwhile, the weaknesses of multiple choices are concluded as follows:

A. The technique tests only recognition knowledge. Namely, the test-taker who can distinguish the correct response without any forms of speaking or writing cannot achieve the purpose of communication and interactions on the basis of the characteristics in the second language learning or acquisition. In some case, there is still a big gap to be bridged between aspects of language knowledge and language performance. This gap will mean that test scores have not provided incomplete information to pupils.

B. Guessing produce a considerable factor. Maybe all of the correct or incorrect responses are the result of guessing.

C. The technique severely restricts what can be tested. The primary issue is that multiple choice items require two or three distractors. The distractors are not always available.

D. Backwash may be harmful. As a vague test point, it is dangerous that practice for the test will have a harmful effect on teaching and learning.

E. Cheating may be promoted. As we all know, children have low self-control ability. The response to a multiple choice test (a, b, c, d) is very easy to communicate to other candidates non-verbally (Hughes, 2009, p78).

The multiple choices of this test paper have taken up 60% in the whole scores. The old criticism of the multiple choice item is as being something that we don't do "in the real word" (Underhill, 1982, cited in Fulcher & Davidson, 2007, p. 63). Therefore, it's necessary to analyse the characteristics of grammatical competence, textual competence, illocutionary competence, and sociolinguistic competence.

4.2.1 The Characteristics of the Length

According to the Syllabus, sentences in Primary Phase mainly based on the simple sentences, do not refer to the complex sentence and the compound-complex sentence, a simple sentence refers to those consisting of only one independent clause: e.g., I love my mother very much. Therefore, the length of the multiple-choice is fit for the curriculum standard.

4.2.2 The Characteristics of the Expected Response

The multiple choice usually consist of four or five/more answers which need the test taker via carefully thinking to choose correctly and the best response for the question. Multiple choice items have many forms, but their basic structure is as follows. There is a stem and a great many options, one of which is correct, while the others are distractors.

4.2.3 The Analysis of the Vocabulary Coverage

In this part, the author has combined with the literature review to analyse the vocabulary coverage. The vocabulary belongs to the part of the grammatical competence. Meanwhile, multiple choice can be conducted to assess the integrated communicative language ability, which includes listening, speaking, reading, and writing etc.

⌘

The proportion of vocabulary in the multiple choice⌘								
Vocabulary	n.	v.	prep.	adv.	art.	adj.	num.	conj.⌘
Number	18	5	0	0	1	3	0	1⌘
Percentage	36%	10%	0%	0%	2%	6%	0%	2%⌘

⌘

Figure 5⌘

From the Figure 5 we can see, the degree of the nouns was accounted for 36% in the Primary English Entrance Test. The following is verbs 10% and adjectives take up 6%. Meanwhile, preposition, adverb and numeral in the whole multiple choices are merely zero. So it can be concluded that this set of paper in the assessing of the young learner' vocabulary existed a situation unbalanced distribution. Judging from the content validity theories and combining with the list of the Syllabus, the test contents were limited, the facets of testing not enough wide-ranging. Therefore, results of the vocabulary coverage are proved with low content validity.

4.2.4 The Analysis of the Grammar Coverage

It's no doubt that the tense must be taken up a large proportion among the communicative language, the tenses and singular & plural are confirmed as the most important grammar point for young children, because he/she needs many grammars to make up appropriate utterance to communicate. Therefore, we can observe from Figure

6 presents tense and past tense which are tested twice and many times in multiple choice. The syllabus required that it's very important to level 2 students to master simple tense. The finding from data can be identified that the difficulty of the task fits for young learners.

The proportion of grammar in the multiple choice ⁶				
Grammar	present tense	Past tense	Singular & plural	There be ⁶
Number	10	11	8	0 ⁶
Percentage	20%	22%	16%	0% ⁶

Figure 6⁶

4.2.5 The Analysis of the Textual Competence

In this part, the author takes the item 8 as an example, mainly from its length, genre, expected response, language structure as analysis.

Length: the passage consists of 5 simple sentences, the most long sentence was made up of 7 words.

Genre: describe a dream

Expected response: selecting a sentence to make a dialogue, rearrange the sentence structure. This aims to make up a complete conversation.

Language structure: Statements, exclamatory sentences, interrogative sentences.

As the Syllabus requirement, the main task of the test is to foster the children's interest and motive to learn English. Combined with the literature review, the author has described the relevant theories on the basis of content validity. Finally, evidences of multiple choice are considered with high content validity.

4.3 The Analysis of the Writing Characteristic

In this section, the part of writing was presented in the last, accounting for 10% in the whole testing.

All in all, the writing part fully presents the spirit of the curriculum and textbooks. To some extent writing has the following characteristics: vocabulary, grammar, function, expression, recognize, emotion, etc. Writing is directly estimated by administer. According to the Requirement of Syllabus, the topic should be close to the young learners' daily life. The genre is supposed to advocate interesting and authentic with some specific purpose of communication. The task is open-ended. Allow young learners to develop his/her own ideas based on the four writing points given.

4.3.1 The Rubric of Writing

A standard test rubric has directly respected to test taker's scores. So the part of the composition is no exception, an excellent item of test writing need to spend enough time or energy to design. On the basis of its open-ended characteristics, different test takers may develop different responses, in terms of differences in individual characteristics and experiences they have. So test designer has to consider all kinds of these factors in processing of testing.

4.3.2 The Analysis of the Testing Introduction

Language: Target language and native language with prompt.

Channel: written

Topic: What a dream!

Requirement: at least five sentences, can develop your imagination appropriately. The first sentence has been given already. Calligraph and format must be beautiful. The sentences must be smoothly, no grammatical errors. Punctuation must be correct.

Bachman (1998) put forward the instruction should be given to the test-taker's mother tongue and target language in order to avoid to being misunderstanding by young learners.

From the intention of the test designer, we can see that the author expected to measure what they wanted to, with the native language help to explain the topic and requirement which can adequately prove his/her expectation, reducing the difficulty of writing tasks. Based on the curriculum standard, primary school students only need to

master preliminary language structures and present tense or past tense. Therefore, the writing task in this paper possesses relevance and representativeness. But there is a little flaw in that the genre in item 8 and item 10 exist a phenomenon which is repeated.

4.3.3 The Analysis of the Scoring Method

In this section, scoring rubric is also listed. The author divided the score into four levels, the score of level 1 has 10 points, level 2 ranged from 7-9 points, level 3 ranged from 1-6 points, and level 4 has no point.

Judging a set of test paper whether or not has high or low content validity to some extent which based on scoring scales. One expected to seek out the scales used in rating performance in the specifications under the heading “criterial levels of performance”. There are two methods to score: holistic and analytic (Vicky, Meg, & Jake, 2009, p. 94).

From the scoring guide we can also see. This paper has the concrete scoring rubric, to some extent reducing some scoring biases, although everything is not perfect. Not every scoring system will give equally valid and reliable results in every situation. To some extent it has proved the content validity of the writing task.

4.3.4 The Analysis of the Expected Response

The expected response in the Framework of the task characteristic includes language, form, and topic characteristics (Bachman & Palmer, 2010). It related to test takers’ response and input for testing materials. So, this study focused on the words, sentences, phrases, context, and grammars. According to the Requirement of the Course Standard, content of expected response matches the young learners’ comprehension and expression.

5. Findings and Recommendations

This design mainly reflected the Requirement of the Primary English Curriculum Standard. After all, it belongs to the low-stake test to face stake-holders, maybe, in many aspects, there are no seriously request like a College Entrance Test. In other words, part of the teachers are less concerned with the results of this test. But, there are still some parents and children to extent that have more concern with it, because it is the best way for them to entrance a good reputation Junior Middle School. To what extent it is still harmful, and causes the negative washback in the process of teaching and learning in Primary School Phase.

5.1 The Strengths of This Test Paper

Firstly, from the characteristic of the test introduction, format, the number of the task, the salience of the task, sequence of the task, relative importance of tasks, time allotment, and recording method we can see that it is familiar to the young learners, also matches the requirement of the Primary English Syllabus. The degree of difficulty of the task fits to the characteristics of the young learner’s age.

5.2 The Weakness of This Test Paper

To begin with, the proportion of subjective items and objective items is not reasonable for communication. The scores of objective items are up to 30 items accounting for 60%. Too many multiple choices belong to restrict the response type, obeying the principle of the language communicative ability. Next, the paper-pencil test is merely written, but it only can assess learners’ listening, reading and writing. The spoken type is still at zero. Face to young learners, the most important task in the language competence is regarded language performance and oral English.

Zangle (2000) claims that the good assessment should reflect the learning process. On the basis of the findings of this paper, the spoken channel seems the best channel for testing young language learners. On the contrary, one of the key reasons of why the written test is proffered over the spoken test. The most important factor might be the task of speaking belongs to direct assessment form, which need to be both time-consuming and a great amount of money or energy to administers (Zangl, 2000). In addition, coverage of grammar and genre/topic, culture, interests were restricted in some aspects, for instance, there be ... etc.

5.3 Recommendations

Therefore, when we are assessing young English language learners, it’s necessary to carry out alternatives in assessment. Such as: the formative and summative assessment, conference assessment, portfolio assessments, self- and peer-assessments, task-based and performance assessment, etc. Oral English must be evaluated in the first place.

6. Conclusions

The purpose of this study was to investigate the content validity of the Junior Middle School English Entrance Test. The content referred to relevant, representitiveness, coverage and weightings based on the data analysis

from the framework of the junior middle school entrance English test. Namely, the analysis of the listening part, multiple choice characteristics and the writing part which include length, genre/topic, question type, expected responses, language competence and scoring method, etc. combined with the test syllabus and the course standard. Finding content of the test paper fits to young learners' practical ability, interesting, and characteristics, it's extremely possible to match the requirement of the Syllabus. By virtue of the task characteristics, it is proved that it has a big relevance between the test contents and the framework of the task characteristics.

By detailed analysis, all the materials are according with young learner's language ability, but still there exist some problems which have not been involved or neglected in testing, such as, cultural awareness, oral testing restricted in a paper and pencil test. In addition, some genres, topic and vocabulary have still presented a phenomenon of redundancy.

Finally, although the author tries her best to make this study more perfect, in terms of limited time and fewer experiences allow this study to exist a little flaw. Investigating the content validity of the junior middle school entrance English testing, some more issues will be involved in this study need deeply to investigate in the future.

Acknowledgement

Thanks Dr. Pan for giving me so much valuable suggestions about the article. In addition, I'm grateful to my families who supported enough time and energies for me to write this paper. Many thanks to them.

References

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation* (pp. 170-196). Cambridge: Cambridge University Press.
- Bachman, L. (1999). *Fundamental Considerations in Language Testing*. Oxford University PR. INC.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in Practice*. Oxford: Oxford University Press.
- Bachman, R. (1998). The factors related to rape reporting behavior and arrest new evidence from the national crime victimization survey. *Criminal Justice and Behavior*, 25(1), 8-29. <https://doi.org/10.1177/0093854898025001002>
- Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice. *Modern Language Journal*, 82(1), 143.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London, NY: Routledge. <https://doi.org/10.4324/9780203449066>
- Fulcher, G., & Davidson, F. (2008). Language testing and assessment: an advanced resource book. *Miscelánea A Journal of English & American Studies*, 283(5), F1098.
- Gronlund, G. (1998). Portfolios as an assessment tool: is collection of work enough? *Young Children*, 53(3), 4-10.
- Henning, G. (1987). *A guide to language testing*. Cambridge Mass: Newbury House.
- Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly, JSTOR*, 24(4), 602-611. <https://doi.org/10.2307/2392366>
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
- Ministry of Education of P. R. C. (2011). *Compulsory Education English Curriculum Standard* (pp. 1-23). Beijing Normal University Press.
- Vicky, Meg, & Jake. (2009). *Testing for Language Teachers*. Cambridge University Press.
- Wainer, H., & Braun, H. I. (1988). *Test validity*.
- Zangle, R. (2000). Monitoring language skills in Austrian primary (elementary) schools: A case study. *Language Testing*, 17, 250-260. <https://doi.org/10.1177/026553220001700208>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).