A Study of the Content Validity of Reading Comprehension Tests in NMET II*

Zhen−Jie Liu & Ming−Hao Jin**

(Dalian Bolun Middle School & Yanbian University)

Liu, Zhen-Jie & Jin, Ming-Hao. (2018). A study of the content validity of reading comprehension tests in NMET II. *The Linguistic Association of Korea Journal, 26*(4), 103-124. Based on Bachman and Palmer's (1996) framework of task characteristics, this paper aims to compare the reading comprehension tests of the 2013-2017 National Matriculation English Test II (NMET II) with the requirements of the New Curriculum Standards and Testing Syllabus, in particular, concerning the content validity of the tests in terms of the characteristics of input and expected responses. The implications of this study are discussed for test designers, teachers, and students. The results show that reading comprehension tests from 2013 to 2017 basically meet the requirements of the New Curriculum Standard and Testing Syllabus and have a high content validity. In terms of the characteristics of input, the length, reading speed, and readability of comprehension passages are appropriate for students; however, the distribution of topics and genres in each year is uneven. In terms of the characteristics of expected responses, the wordings of questions and the design of question items are reasonable. However, the response type is mainly restricted to selected response questions (multiple-choice), and the measure of reading skills coverage is unbalanced.

Key Words: content validity, reading comprehension test, NMET II

1. Introduction

Language tests, especially large-scale tests may produce tremendous influence on language teaching and learning. A test will influence teaching, learning, what and how teachers teach, and teachers' attitudes to the content and method of teaching.

The National Matriculation English Test is a large-scale nationwide talent selection test. Due to its high washback on English teaching and learning, the quality of the test paper is of great importance. Reading comprehension scores account for 26% of total scores, so it's a very important part of the NMET. Up till now, the provinces that use the NMET paper II for college entrance examination include Gansu, Qinghai, Inner Mongolia, Heilongjiang, Jilin, Liaoning, Ningxia, Xinjiang, Tibet, Shaanxi, Chongqing, and Hainan Provinces. However, there has been very little research done on content validity or design quality of the NMET paper II.

This paper aims at exploring the content validity of reading comprehension in the NMET II from 2013 to 2017, analyzing whether the characteristics of input and expected response as proposed by Bachman & Palmer (1996) are in line with the requests of the New Curriculum Standard and Testing Syllabus. This paper will also provide some beneficial information and suggestions for designing test papers, teaching methods and enhancing students learning. Hopefully, this study provides some useful feedback on the reading comprehension section by putting forward some suggestions regarding reading length, topic, genre, readability, etc. It is beneficial for teachers to better understand students' requests about how to improve their reading ability so that teachers can guide students to learn effectively and students may become much clearer about their targets in doing reading comprehension.

2. Literature Review

2.1. Basic Concepts

2.1.1. Content Validity

Bachman (1990) insisted that the content validity can be investigated from two aspects: content relevance and content coverage. Content relevance refers to the consistency between tested content and those teaching syllabus, testing syllabus and testing specifications. Content coverage is the extent to which the tasks designed in a test adequately represent the behavioral domain in question. Alderson, et. al. (2000) also agreed that a common way to guarantee the content validity is to compare test content with the teaching and testing syllabus, and to see whether the test measures what they should be. Wood (2001) believed content validity should adhere to a blueprint, which relates to official documents, such as the teaching and testing syllabus.

The content validity in this study refers to the extent to which the NMET II Reading Comprehension Test from 2013 to 2017 is consistent with the New Curriculum Standard and the Testing Syllabus.

2.1.2. Reading and Reading Comprehension Test

Widdowson (1978) thought reading is a process of communication between writer, text, and readers, in which reading and understanding is a two-way interaction. Writers write passages to convey their own ideas. For readers, reading is a way to understand what a writer's intention is and as well as a tool to improve their reading comprehension.

Weir (1993) illustrated that reading is a process of interacting and choosing what happens between the reader and reading materials. The readers' background knowledge and language knowledge will communicate with information in the reading materials, which leads to their understanding.

According to different purposes of reading comprehension tests, different test designers adopt different test formats. Some typical answers are shown in the following:

(1) Multiple-choice Question. At present, this is one of the most popular test formats in large public examinations, such as TOEFL, NMET, CET 4/6 (College

English Test Level 4/6), TEM 4/8(Test for English Majors Level 4/8) and so on. It is used in nearly every large-scale examination as an objective item and has many advantages. Firstly, the scores are calculated by computer or machine, which guarantees their objectivity and reliability. Secondly, it's fast, saves time and effort, and what's more, it is easy to analyze answers so that efficient feedback is provided on actual classroom teaching. However, a drawback is that it is difficult to assess whether the answers made by test takers are from guessing. Therefore, test takers' real competence can't be tested directly.

(2) Short-answer Question. These questions require that the test taker provide short, multi-word responses so that the possibility of guessing is greatly reduced. For test takers, a comprehensive understanding of reading materials is needed. At the same time, it requires the test takers to engage their writing skills, which tests their output competence rather than recognition competence. As opposed to Multiple Choice, the scoring of Short-Answer Questions is done by humans and therefore introduces the possibility of bias, which reduces the reliability of scores.

(3) True-or-False Question. These are widely used by test designers to test reading competence because it's easy to judge whether the answer is right or wrong. However, its disadvantages are obvious. Test takers have a 50 percent probability to get the right answer even if they can't understand the reading materials, which makes it difficult to distinguish a test taker's real language competence.

However, it is suggested that in order to improve test reliability, all test items in the reading comprehension section should be Multiple-choice questions.

2.1.3. New Curriculum Standard

Curriculum Standards for Full-time Compulsory Education and Senior High Schools (Trial Version) were issued by the Ministry of Education of the People's Republic of China in 2001. Thereafter, in line with the requirements of the senior high school curriculum reform, the English Curriculum Standard for High School Students (abbreviated as *the New Curriculum Standard*) was issued in 2003. It was determined that the Senior High School students' reading ability should reach level 8 when they graduate.

2.1.4. Testing Syllabus

In order to instruct the NMET and define the content and requirements of a test, the Test Center of the State Ministry of Education drafts a Testing Syllabus every year, which is according to the Institutions of Higher Education's requirements for students' knowledge and competence. It is made according to the New Curriculum Standard and actual classroom teaching situations, and the content each year is similar.

2.2. Theoretical Basis

In 1990, Bachman put forward the Framework of Task Characteristics for the first time. It can help people describe present tests, design a test, and testify the validity of a test. What's more, it can be used as the basis of hypotheses in language test research (Qiu, 2010). On the basis of his theory, Bachman together with Palmer put forward a more specific framework of task characteristics in 1996, and formed a relatively complete theory. They believe language use task is an activity that involves individuals in using language for the purpose of achieving a particular goal in a particular situation. This framework consists of the following five parts: characteristics of the setting, characteristics of the test rubrics, characteristics of the input, characteristics of the expected response and relationship between input and response. Among these five characteristics, input and expected responses have been chosen to analyze the content of validity in this paper.

2.3. Previous Studies Abroad and at Home

The study of reading comprehension abroad mainly focuses on the variables and test format which influences reading comprehension, and factors influencing the level of difficulty in the reading comprehension materials, etc.

In terms of the variables influencing reading comprehension, Alderson (2000) divided variables into two types: One is the reader variable, the other is the text variable. The reader variable refers to a reader's knowledge, motivation, purpose and other characteristics, while text variable includes the topic, content, type, genre and so forth. This research focuses on the reading comprehension test in the NMET II from 2013 to 2017, in particular the text variables.

With regard to the test format of influencing reading comprehension, different scholars have their own opinions. Samson (1983) performed an experiment which revealed that different test tasks didn't influence test takers' performance, but many scholars doubted Samson's research, therefore they did some related experiments. Schohamy (1984) and Kobayashi (2002) pointed out that different test tasks did indeed influence test takers' performance.

Based on the above research, the test format should be taken into account when analyzing the content validity of reading comprehension tests.

Nowadays the studies of reading comprehension in the NMET in China are mainly divided into two parts, the studies of national test papers and test papers designed by autonomous proposition provinces.

2.3.1. Research on National Matriculation English Test Paper

Dong (2010) analyzed the content validity of 75 passages and 300 questions in the National Matriculation English Test from 2003 to 2009 according to characteristics of input and expected response. The results show that the length of passages and reading speed are appropriate, and the genre and topic are extensive. So it basically meets the requirements of the New Curriculum Standard and Testing Syllabus, but the reading speed requirement in test paper is low. What's more, the measure of reading skills coverage is unbalanced and comprehensive.

Tian (2012) studied every part of the NMET I used by Shanxi Province after the new curriculum reform in 2011, and compared the collected data with the New Curriculum Standard, Testing Syllabus and Testing Specifications. His study revealed that it has a high content validity and reflects the measure of students' language competence.

Liu (2016) also studied the content validity of the NMET I, but she did a diachronic research from 2012 to 2015 according to five aspects, which included characteristics of the setting, characteristics of reading tests rubrics, characteristics of input, characteristics of expected response, and relationship between input and expected response. It proposes that the structure and pattern of test papers are stable, the length of passage is appropriate, and the topic is extensive and interesting. What's more, six reading skills stipulated in the testing syllabus are all covered.

2.3.2. Research on the MET paper designed by autonomous proposition provinces

Zhang (2013) studied the reading comprehension's content validity of 2007-2012 Shandong Matriculation English Test papers by comparing it with the New Curriculum Standard, Testing Syllabus and Teaching Syllabus. The result shows that it has a high content validity.

Xiao & Lin (2015) analyzed the content validity of reading comprehension in 2010-2014 Hunan Matriculation English Test Papers. In order to find out whether the reading comprehension tests meet the command of the New Curriculum Standard and Testing Syllabus, they analyzed the aspects of length, reading speed, proportion of new words, readability, genre and reading skills measured. The study suggested that the order of passages should be redesigned more reasonably.

Hu & Wei (2016) did some research on the content validity of reading comprehension tests in 2011-2015 Hubei Matriculation English Test papers, and they put forward that English teachers should increase reading exercises of practical writing and focus more on passage structures and attitude.

To sum up, the studies on content validity of reading comprehension tests in the NMET have increased gradually in recent years, but most of them focus on analysis of test papers in autonomous proposition provinces, and less attention has been given to the study of content validity of the NMET paper II, which test is taken by students from another 11 provinces in China, like Jilin Province.

3. Research Design

In this study, Bachman & Palmer's framework of task characteristics (1996) is used to analyze the content validity of the NMET II from two aspects, characteristics of input and expected responses, shown in Table 1.

Jilin Province has begun to use the NMET II, by the instruction of New English Curriculum Standard since 2013. Consequently, this study takes all 20 reading passages and 75 question items in the NMET II reading comprehension tests from 2013 to 2017 as its sample.

This study aims to investigate the content validity of reading comprehension tests in 2013-2017 NMET paper II, and the research questions are as follows:

(1) To what extent do the characteristics of input in 2013-2017 NMET paper II correspond with the New Curriculum Standard and Testing Syllabus?

(2) To what extent do the characteristics of expected response in 2013-2017 NMET paper II correspond with the New Curriculum Standard and Testing Syllabus?

(3) By doing this research, what are the suggestions on developing students' reading comprehension skills and competence?

Table 1. Testifying Indicators of 2013-2017 NMET II Reading Comprehension

Characteristics of Input	Length	passage only, passage and question items
	Topic	24 topics stipulated in New Curriculum Standard
	Genre	narration, exposition, argumentation, practical writing
	Reading Speed	each passage, entire reading comprehension
	Readability	Flesch Reading Ease
Characteristics of Expected Response	Response Type	selected response limited Production response extended Production response
	Reading Skills Coverage	(1) understanding the main idea of the passage (2) understanding specific information (3) inferring new words and phrases through context (4) making simple judgement and inference (5) understanding the basic structure of passages (6) understanding the author's attention and attitude
	Question Items	language, designing

After data collection, this paper analyzes collected data and compares it with the New English Curriculum Standard and Testing Syllabus, which are official documents to provide instructions for teachers and test designers. In the process of analyzing the data, Microsoft Office Word 2007 is used to calculate the numbers of words in passages, reading speed, as well as tables and graphs that makes the reading comprehension data more direct and vivid. The readability of 20 reading passages in this study (question items excluded) is calculated by Microsoft Office Word 2007, which has a tool to calculate Flesch Reading Ease.

4. Results and Discussion

4.1. Analysis on Characteristics of Input

Characteristics of input are analyzed in terms of five aspects: length, topic, genre, reading speed and readability.

4.1.1. Length

According to the requirements of the New Curriculum Standard, the number of reading passages in a test paper should be 3 to 5, and the total words of reading passages should be 300-1500. That means each passage may have 100-300 words. What's more, the Testing Syllabus issued every year stipulates that the total words in reading comprehension test should be more than 900 words, which means that the average length of each passage should be above 225, as shown in Table 2.

Table 2. Text Length of Reading Comprehension Tests in 2013-2017 NMET II

Year	PA	PB	PC	PD	AL	TWP	TWQ	TW
2013	295	261	208	242	252	1006	514	1520
2014	285	270	285	265	276	1105	471	1576
2015	260	322	263	269	279	1114	462	1576
2016	285	306	252	280	281	1123	463	1586
2017	302	304	326	292	306	1224	471	1695

Note: P=Passage; TW=Total words; TWP=Total words of passage; TWQ=Total words of questions; AL=Average length

Table 2 shows that the total words of the reading comprehension test meet the requirements of the Testing Syllabus. The length of each passage (except the words of question items) in 2013-2017 NMET paper II, which ranges from 208 to 326, is quite stable, and the average passage length of each year is more than 225 words. In this regard, the length of each passage is in line with the requirements of the New Curriculum Standard. However, the text length distribution should follow the principle of short to long and simple to complex.

4.1.2. Topic

In the New Curriculum Standard, 24 topics are stipulated as follows: Personal

information; Family, Friends and people around; Personal environments; Daily routines; School life; Interests and hobbies; Emotions; Interpersonal relationships; Plans and intentions; Festivals, holidays and celebrations; Shopping; Food and drink; Health; Weather; Entertainment and sports; Travel and transport; Language learning; Nature; The world and and the environment; Popular science and modern technology; Topical issues; History and geography; Society; Literature and art, as shown in Table 3.

Table 3. List of Topics in 2013-2017 NMET II

Passage	Topic	Types of topic
2013(A)	Be brave to tell your ideas	E&E
2013(B)	Brief introduction to international festival of art in Edinburgh	S&C
2013(C)	Singapore's Chocolate Research Facility	S&C
2013(D)	Several low-cost gifts for Mother's Day	S&C
2014(A)	Have faith and trust in people	E&E
2014(B)	America's achievements in environment protection	H&N
2014(C)	The latest trend in American Childcare	S&C
2014(D)	Metro pocket guide	S&C
2015(A)	The bother a new TV set brought to the author	E&E
2015(B)	The effects on figures caused by house's environment	S&C
2015(C)	The phenomenon of gap year	S&C
2015(D)	4 programs of one-day tour	S&C
2016(A)	Art activities related to music and drama	S&C
2016(B)	Author's experience of encouraging students' creativity	E&E
2016(C)	The website BookCrossing.com	S&C
2016(D)	A collection of photos brings an unsuccessful Antarctic voyage back to life	S&C
2017(A)	4 kinds of theaters	S&C
2017(B)	Sincere friendship	E&E
2017(C)	Flying Car	S&T
2017(D)	Leafy plant will give out VOCs when it's under attack	H&N

On the basis of the Curriculum Standard, the Testing Syllabus, and reading passages in the NMET paper II, topic can be classified into 4 types as follows: Social life and Culture (S&C), Experience and Emotion (E&E), Science and Technology (S&T), and Human and Nature (H&N), as shown in Table 4.

Table 4. Topic Analysis of Reading Comprehension in 2013-2017 NMET II

	Social life and Culture	Experience and Emotion	Science and Technology	Human and Nature
2013	3	1	0	0
2014	2	1	0	1
2015	3	1	0	0
2016	3	1	0	0
2017	1	1	1	1
Total	12	5	1	2
Rate	60%	25%	5%	10%

It can be seen from Table 3 and Table 4 that the distribution of topic in each year is narrow and uneven, especially in 2013, 2015 and 2016. The topics are severely uneven. And passages about "social life and culture" make up quite a large proportion in the past five years, even up to 60%. Passages about experience and emotion account for 25%, while passages relating to science and technology only take up 5%. Therefore, it is obvious that test designers of the NMET II reading comprehension focus more on topics that are closely related to test takers' daily life. What's more, many selected topics are related to moral education, such as passage A and D in 2013, passage A in 2014, passage B in 2016, and passage B in 2017.

4.1.3. Genre

Four genres are analyzed in the study, including narration, exposition, argumentation and practical writing, as shown in Table 5.

Table 5. Genres of Reading Comprehension in 2013−2017 NMET Ⅱ

	Narration	Exposition	Argumentative	Practical writing
2013	1 (A)	2 (BC)		1 (D)
2014	1 (A)	2 (BC)		1 (D)
2015	1 (A)	1 (B)	1 (C)	1 (D)
2016	2 (BD)	1 (C)		1 (A)
2017	1 (B)	2 (CD)		1 (A)
Rate	30%	40%	5%	25%

Different genres have different features, therefore the requirements of readers' reading skills in different genres are quite different: 1) Narration genre focuses on narrating characters, incidents, scenes and objects, the basic content

of which is the writing characters' experiences and changes of things. For test takers, it is relatively easy to understand; 2) Exposition genre explains the features of things by presenting plenty of information, which is a little bit more difficult than narration; 3) The purpose of argumentation genre is to demonstrate a certain opinion right or false by giving some facts, which needs logical thinking, critical thinking and the ability to make inferences; 4) Practical writing concentrates on practicability, reality, timeliness and normalization. Reading comprehension mainly checks test takers' ability to find and deal with information, which includes not only selecting and comparing specific details, but also judgement and inferences.

Table 5 implies that the distribution of reading comprehension in 2013-2017 NMET paper II was not balanced. Exposition genre takes a large proportion accounting for 40%, while argumentation genre takes only 5%, which is extremely unbalance. But the arrangement order of genre in 2013-2017 NMET paper II is reasonable. Narration genre is almost arranged as in the former part of reading comprehension, which closely relates to test takers' daily life, and it is helpful to reduce anxiety and stress so that they can concentrate on the test itself.

4.1.4. Reading Speed

In the specifications of Testing Syllabus, it is stipulated that the reading comprehension should be accomplished within 35 minutes. Therefore, the reading speed can be calculated by total words of readings. However, the time for passages and question items are not the same.

Yang & Weir (1998) stated that the time of passage and the time of question items are about 1 to 0.75. That means, within 35 minutes, students have 20 minutes to read passages, and 15 minutes to read the question items. The reading comprehension test of 2013-2017 NMET paper II has two parts. The first part is reading with multiple-choice questions and the second part is reading to complete the passage (select 5 from 7 sentences given). But this study analyzed 5 reading passages with multiple choice questions, therefore the time to read passages is 16 minutes, and the time to read question items is 12 minutes. In this study, according to proportion of 16:12, only the reading speed of passages and the total input are going to be calculated respectively, as shown in Table 6.

Table 6. Reading Speed of Reading Comprehension in NMET II

	Total words of passages	Minutes	Words per minute	Total words including questions	Minutes	Words per minute
2013	1006	16	62.9	1520	28	54.3
2014	1105	16	69.1	1576	28	56.3
2015	1114	16	69.6	1576	28	56.3
2016	1123	16	70.2	1586	28	56.6
2017	1224	16	76.5	1695	28	60.5

The teaching syllabus describes 2 levels of reading speed. The lower level demands 50-70 words per minute, and the higher level demands 70-80 words per minute. From Table 6, it can be seen that the reading speed increases year by year from 2013 to 2017, which means the demand of test takers' reading speed improves gradually, and the reading speed of reading comprehension in 2013-2017 NMET paper II is designed in accordance with requirements of the teaching syllabus, and therefore it has a high content validity.

4.1.5. Readability

There are some formula to assess readability, such as the SMOG Index, the Fry Readability Formula, and the Flesch Index. Among them, the Flesch Index is the most widely used. The Flesch formula produces a Flesch Reading Ease (abbreviated as FRE) score: $FRE = 206.835 - 0.846wl - 1.015sl$.(quoted in Alderson, 2000). "wl" refers to the number of syllables per 100 words, and "sl" refers to the average number of words per sentence.

Table 7. Reference Table of Flesch Reading Ease

0-30	Very difficult
30-50	Difficult
50-60	Fairly difficult
60-70	Standard
70-80	Fairly easy
80-90	Easy
90-100	Very easy

The score mapping table of the Flesch Reading Ease is shown in Table 7. The readability of 20 reading passages in this study (question items excluded) is

calculated by Microsoft Word 2007, which has a tool to calculate Flesch Reading Ease, as shown in Table 8.

Table 8. Readability of Reading Comprehension in 2013-2017 NMET II

	Passage A	Passage B	Passage C	Passage D	Average
2013	71.1	41.9	44.2	75.2	58.1
2014	72.2	50.3	59.6	43.6	56.4
2015	53.6	80.9	43.2	56.2	58.5
2016	75.5	77.9	76.2	45.7	68.8
2017	44.4	66.7	54.9	62.1	57.0

Table 8 shows that the average readability of reading comprehension in 2013-2017 NMET paper II ranges from 56.4 to 68.8, which assesses as "fairly difficult", "standard", and "fairly easy" according to the Table 7. The NMET is a selective test, and it meets to the requirements of the NMET reading materials which should have certain difficulty and basic discrimination so as to pick up talents for higher education. However, the readability distribution is not reasonable. In the year of 2015 and 2017, the readability of passage A relates to "fairly difficult" and "difficult", which may increase the anxiety of test takers. The following table shows the general rate of readability of the total 20 reading passages in the NMET II from 2013 to 2017, as shown in Table 9.

Table 9. Rate of Readability

	Numbers of passages	Rate
Very difficult	0	0
Difficult	6	30%
Fairly difficult	5	25%
Standard	2	10%
Fairly easy	6	30%
Easy	1	5%
Very easy	0	0

In specific, "difficult" and "fairly difficult" accounts for 55%, but "standard" and "fairly easy" account for 40%. It shows that the reading comprehension test of the NMET II paper is reasonable because it not only takes test takers' language level into account, but also designs some passages which are difficult

to read.

4.2. Analysis on Characteristics of Expected Response

Three aspects are going to be covered in this section, including response types, reading skills coverage, and question items.

4.2.1. Response Types

According to *Language Testing in Practice* written by Bachman and Palmer in 1999, response types are classified into three, namely: Selected response, limited production response and extended production response. Selected response measures test takers' reading competence by giving some options, which is represented by Multiple-choice questions. Limited production response requires test takers to write several words, phrases or a sentence, such as the Short Answer questions. In terms of extended response, this is a free writing, which requires test takers to write several sentences.

The response type of reading comprehension in 2013-2017 NMET paper II relates to selected response, which means Multiple-choice was the only test format used. There is no doubt that it ensures high objectivity and reliability, and what's more, it is convenient and efficient. However, the process of answering Multiple-choice questions is not communicative or interactive, and it can't measure test takers' real language ability and comprehensive competence. So it is suggested that various kinds of test formats should be added, such as Short-answer questions and True-or-False questions, to better examine test takers' reading ability.

4.2.2. Reading Skills Coverage

Alderson (2000) indicated that the reading performance of test takers was affected by a series of reading skills, reading competence and reading strategies.
There are 6 types of reading skills stipulated in theTesting Syllabus:
(1) Understanding the main idea of the passage
(2) Understanding specific information
(3) Inferring new words and phrases through context
(4) Making simple judgement and inference

(5) Understanding the basic structure of passages

(6) Understanding the author's attention and attitude

There were 75 question items in total in 2013-2017 NMET paper II. The reading skills' coverage in the NMET II reading comprehension test was counted according to the Testing Syllabus, as shown in Table 10.

Table 10. Reading Skills Coverage of Reading Comprehension in NMET II

	Main Ideas	Specific Information	Deducing Lexical Items	Judgement and Inference	Structure of Passages	Attention and Attitude
2013	1	11	2	1	0	0
2014	1	11	2	1	0	0
2015	1	10	1	2	0	1
2016	1	8	2	3	0	1
2017	2	8	2	2	0	1
Total	6	46	9	11	0	3
Rate	8%	61%	12%	15%	0	4%

Table 10 shows that the distribution of reading skills coverage is extremely unbalanced. No question item was set to measure the structure of passages in those five years. What's more, in 2013, 2014 and 2016, structure of passage, attention and attitude, the measure of these two skills were excluded.

On the other hand, the reading comprehension test in the NMET paper II focuses on test takers' basic reading skills. 46 questions were set to measure specific information in 75 question items, which accounts for 61%. The measure of understanding main ideas is only 8%, and the measure of making judgement and inference accounts for 15%. Understanding specific information is relatively easy, which belongs to lower reading skill and demands test takers to locate the detailed information and compare it with options. Understanding main ideas and making judgement and inference belongs to higher reading skills, which requires test takers' comprehensive understanding of a passage and logical thinking.

The measure of higher reading skills only takes up 23%, while the measure of lower reading skills accounts for 61%, which demonstrates that reading comprehension tests in the past five years concentrated more on measuring test takers' basic reading skills. It may result in the controlled items made by the

examiners did not correctly discriminate the high level students as proposed by Kim, Nanu & Lee, Heechul (2017).

4.2.3. Question Items

(1) Wordings of question items

Alderson (2000) suggested that if the wordings of questions are more difficult to understand than the text itself, we can't tell whether it is the difficulty of passages or questions that causes poor performance. According to the analysis on words used in question items of the reading comprehension test, it shows that the number of words in question items is appropriate, and the words used in question items are mostly common and easy for test takers to understand. Simple sentence pattern is adopted, which basically follows the structure of <Subject+Predicate+Object>. Therefore, the wordings of question items are appropriate.

(2) Designing of question items

On one hand, the expression of some question items has similarities and is consistent. Considering that some question items are text-oriented, only the features of the following three readings skills are going to be analyzed.

The reading skill of **understanding the main idea of the passage**:

Which of the following can be the best title for the text? (2013, A/2014, A)

What can be a suitable title for the text? (2015, B)

What is the best title for the text? (2016,C/2017,C)

What is the first paragraph mainly about? (2017,C)

The reading skill of **deducing unfamiliar lexical items through context**:

In the... paragraph... probably means_____. (2013, A)

The word... in paragraph... probably refer to..._____. (2013, C)

What does the term "..." in the text mean? (2014, C)

Which of the following can best replace the phrase...in paragraph..? (2015, A)

What does the underlined word...in paragraph...probably mean/refer to? (2014, A/ 2016, B/ 2016 , C/ 2017, B)

The reading skill of **understanding the author's attention and attitude**:

How does the author sound when telling the story? (2015, A)

Why does the author mention... in the... paragraph? (2016, C)

What is the author's purpose in writing the text? (2017, B)

According to the statements above, it can be noted that the expressions of a certain questions in the different years are exactly the same. It is worth reference for reading comprehension test designers.

On the other hand, in terms of designing stems and options, Heaton (2000) illustrated that question stems could be shown in three forms as follows: An incomplete statement, a question, and a complete statement. What's more, the tested points should distribute to each part of the whole passage, not only the front part or the last part. It is found that question stems are reasonable, and that the distribution of tested points is appropriate in all 20 reading passages.

5. Conclusion and Implication

5.1. Major Findings

5.1.1. In terms of characteristics of input:

The total text length of reading comprehension test meets the requirements of the Testing Syllabus. The length of each passage is quite stable, which ranges from 208 to 326 words. However, the text length distribution can be arranged more scientifically. The distribution of topics is narrow and uneven. Reading comprehension test designers focus more on topics that are closely related to test takers' daily life. What's more, many selected topics are related to moral education. The genre distribution of reading comprehension is unbalanced. The reading speed increased year by year from 2013 to 2017, and it means the demanding of test takers' reading speed has increased gradually. It is found that the reading comprehension test is designed in accordance with requirements of the teaching syllabus and also has a high content validity. The average readability of reading comprehension in 2013-2017 NMET paper II meets the requirements of the NMET. However, the readability distribution order of each year could be adjusted.

5.1.2. In terms of characteristics of expected response:

Multiple-choice question was the only test format in reading comprehension tests of 2013-2017 NMET II, so various test formats should be taken into

consideration. The reading skills coverage was extremely unbalanced. Concerning the language of question items, the number of words in question items is appropriate. The words in question items were easy for test takers to understand. Concerning the designing of question items, the expression of some question items has similarities and consistencies.

5.2. Implication

5.2.1. Implications for test designers

The arrangement order of passages in reading comprehension tests should be more reasonable, following the principle of short to long, simple to complicated, which would be helpful to reduce test takers' anxiety and pressure. Test designers should select a variety of topics which are enriching and diversified in order to make sure the extensiveness of reading passage not just limited to "social life and culture", and "experience and emotion". The genre distribution of reading comprehension should be balanced, and there should be an increase on the proportion of argumentation genre, such as invitations, notices, book review and so on. Various test formats, such as True-or-False question and Short-answer question, should be taken into consideration in order to measure test takers' comprehensive language competence. Regarding the reading skill coverage, the measure of "structure of passage" and "attention and attitude" should be included.

5.2.2. Implications for English teaching and learning

English teachers and students should pay more attention to the washback effect of testing. The New Curriculum Standard and the Testing Syllabus are very important guidelines for English testing. So it is necessary for English teachers and students to study the direction of the NMET to be better prepare for the examination. It is suggested that more reading materials related to students' daily life, hot topic news and other extracurricular reading materials with different kinds of topics are used to broaden their horizons, enlarge their vocabulary and grasp more background information. It is also helpful to instruct students to have a good understanding of four kinds of genres, such as features of language, structure and topics. Students should improve their reading speed

and develop a good habit of reading often. English teachers can request students look through the reading materials in a limited time, and inform students of time left so that they can improve their reading speed. What's more, teachers should put emphasis on training students' reading skills, and make full use of modern teaching technology. It is suggested to avoid abstract explanations, but rather to find appropriate methods which make it easier for students to understand and apply.

References

Alderson, J.C., Clapham, C., & Wall, D. (2000). *Language test construction and evaluation*. Beijing: Foreign Language Teaching and Research Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Shanghai: Shanghai Foreign Language Teaching Press.

Dong, M. X. (2010). Diachronic analysis and research on the reading comprehension test of National Matriculation English Test. *Foreign Language Teaching in Primary and Secondary Schools (Middle School)*, 2, 31-37.

Hu, B., & Wei, J. S. (2016). Research on content validity and difficulty of English reading comprehension of Matriculation English Test in Hubei. *Overseas English, 15*, 91-93.

Kim, N., & Lee, H. (2017). An exploration of high school reading test items based on discriminant analysis. *The Linguistic Association of Korea Journal, 25*(1), 75-92.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19*, 193-220.

Liu, S. Y. (2016). A study of content validity on reading comprehension tests of NMET (2015). Unpublished master's thesis, Jiangxi Normal University, Nanchang City, Jiangxi Province, China.

Qiu, J. (2010). Content validity study on reading comprehension tests of NMET of Chongqing. Unpublished master's thesis, Southwest University, Chongqing City, Chongqing Municipality, China.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1*(2), 147-170.

The Ministry of Education of the People's Republic of China. (2013). Curriculum standards for full-time compulsory education and senior high schools (Trial Version). Beijing: People's Education Press.

Tian, X. L. (2012). The content validity on the National Matriculation English Test (NMET) of the year 2011 guided by the new curriculum standard. Unpublished master's thesis, Shanxi Normal University, Xi'an City, Shanxi Province, China.

Weir, C. J. (1993). *Understanding and developing language tests*. Hemel Hemstead: Prentice Hall International (UK) Ltd.

Widdowson, H. G. (1978). *Teaching language as communication.* Oxford: Oxford University Press.

Wood, R. B. (2001). *Assessment and testing: A survey of research.* Beijing: Foreign Language Teaching and Research Press.

Xiao,Y. N., & Lin, C. (2015). A study on content validity of reading comprehension test of Matriculation English Test in Hunan (2010-2014). *Education Measurement and Evaluation, 3*, 44-49.

Zhang, X. H. (2013). Content validity study on reading expression tests of NMET (Shandong). Unpublished master's thesis, Shandong Normal University, Jinan City, Shandong Province, China.

Liu, Zhen-Jie
Junior middle school English teacher
Dalian Bolun Middle School
58# Quanshui H3 Community, Ganjingzi District, Dalian City, Liaoning Province, China
E-mail: lzj_712@163.com

Jin, Ming-Hao
Associate Professor, Chair of English Department
Department of English, Yanbian University
977 # Gongyuan Road, Yanji City, Jilin Province, China
E-mail: jinminghao@ybu.edu.cn