English Language & Literature Teaching, Vol. 19, No. 4 Winter 2013

Content Validity of Language Knowledge Sections of Four National Matriculation English Test Papers in China

Ming-Hao Jin*
(Yanbian University)
Byoung-Man Jeon**
(Chonbuk National University)

Jin, Ming-Hao & Jeon, Byoung-Man. (2013). Content validity of language knowledge sections of four national matriculation English test papers in China. *English Language & Literature Teaching, 19*(4), 31-53.

This study attempts to analyze the content validity of the language knowledge sections of four National Matriculation English Test (NMET) papers administered in 2012 from the perspective of the New Senior English Curriculum Standards (NSECS) and the 2012 test syllabus in order to see to what extent the language knowledge sections of the NMET meet the requirements of the NSECS and specifications of the test syllabus. The results show that all items of the language knowledge sections of four NMET papers had relatively high content validity with its design corresponding to the requirements of the NSECS and the 2012 test syllabus in terms of basic grammar and grammatical use of vocabulary. However, some deficiencies were found which might affect the content validity of the NMET. Most test items of the language knowledge sections were multiple-choice questions so that the expected responses were limited to selected production responses. It is suggested that limited and extended production responses should be increased in order to improve quality of the language knowledge section. The types of test items should also be diversified to examine test takers' higher level of language knowledge.

[content validity/language knowledge/the National Matriculation English Test/the New Senior English Curriculum Standards]

* First author, ** Corresponding author

교보문고 KYOBO Book Centre

I. INTRODUCTION

English has been regarded as one of the most important communicative tools around the world and English education plays an important role in social and cultural education in China. Although China has made great progress in various aspects of theoretical studies such as teaching methodology, curriculum design, and textbook development, there are still several areas of improvement to reform English education. Among many factors, testing has a vital function in English education system, which has been in place for decades. However, as one of the most important factors of validity in a test, content validity has not been given much attention, especially at the time of implementing the New English Curriculum Standards (Experimental manuscript) for ordinary secondary schools for ten years (a revised one in 2003), including the New Primary English Curriculum Standards for primary schools, the New Junior English Curriculum Standards for junior middle schools, and the New Senior English Curriculum Standards for senior high schools.

Since the National Matriculation English Test (NMET) is a large scale high-stake test, which is supposed to exert profound influence on English teaching and learning with its content and format at the secondary schools (Jeon, 2004; Madaus, 1988), it is vital that the test has a positive washback effects. Thus, it is urgently needed to investigate the content validity of the NMET, through evaluating what the qualities of the test items are, to what extent the tests are reliable and valid, and whether the tests adequately measure the test takers' language ability to meet the requirements of the New Senior English Curriculum Standards (NSECS) for high schools.

Unfortunately, there are few researches on the content validity of the NMET, just limited to Gu and his research team's series of studies on the content validity of writing section of the 2007 NMET papers (Gu & Gao, 2007), listening section of the 2008 NMET papers (Gu & Lei, 2009) and writing section of the 2009 NMET papers (Gu & Yang, 2010) in terms of content and design from the perspective of the NSECS. Another research was conducted by Dong (2011) on the content validity of reading section of the Chongqing NMET papers from 2004 to 2009. All these studies pointed out that the NMET has a relatively high content validity, but there is still much room for improvement. However, no study was conducted to see the content validity of the language knowledge section of the NMET and thus the present study investigates the language knowledge sections of four NMET papers in terms of quality and consistence.

The fundamental goal of the NSECS is to develop students' comprehensive English language competence. It is used to promote quality-oriented education as opposed to exam-oriented education, task-based teaching methods, and experiential and participatory learning.

교보문고 KYOBO Book Centre

Hence, the purpose of this study is to discuss the content validity of four representative NMET papers administered in 2012 to find out whether those test papers have met the requirements of the NSECS (the NMET version of NEC, Beijing, Shanghai and Guangdong)

II. THEORETICAL BACKGROUND

A good test should have high quality so that it can measure learners' language skills accurately and show off teaching contents and results more effectively. A practical and authentic test with high level of reliability and validity can improve students' learning attitudes, motivation, and learners' performance, and even have positive influence on language teaching and learning. The most important consideration in designing a language test is its usefulness or principles-practicality, reliability, validity, authenticity, and washback (Brown, 2004).

As an important aspect of testing and measuring, validity has been defined in many ways. American Psychological Association (1985) defined validity as "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (p. 9). Messick (1989) claimed that validity is "an integrated evaluative judgment of the degree to which empirical evidences and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p. 13). In simple words, validity shows the relationship between evidence from test performance and inferences about test-takers abilities (McNamara, 2000). Furthermore, Brown (2004) defined validity as "the degree to which a test measures what it is supposed to measure, or can be used successfully for the purpose for which it is intended" (p. 22).

From the above definitions of validity, it can be seen that validity has been studied over the years by different language testing experts from various perspectives, which results in different terms or new terminologies to account for the same concept (McNamara, 2000; Messick, 1989; Weir, 1990), and makes the study of validity more complicated and confusing than perhaps it needs to be. Among those definitions, the one proposed by Brown (2004) is widely accepted. According to him, validity can be divided into five categories, content validity, criterion-related validity, construct validity, consequential validity, and face validity.

교보문고 KYOBO Book Centre

1. Content Validity

Oller (1979) pointed out that content validity is a form of validity that is based on the degree to which a test adequately and sufficiently measures the particular skills or behavior it sets out to measure. A test is said to have content validity if its content constitutes a representative sample of the language skills, structure, etc. with which it is meant to be concerned (Hughes, 2003, p. 26). Henning (1987) claimed that content validity is concerned with whether or not the content of the test is sufficiently representative and comprehensive for the test to be a valid measure of what it is supposed to measure. It can be seen that content validity is of great importance in that it can influence construct validity and washback of a test (Hughes, 2003).

The most concern in content validity is to be judged by test items of test paper whether they cover what are required in the teaching syllabus or in the testing specifications because the teaching syllabus or the testing specifications usually provide detailed descriptions of what should be tested in a test. Content validity depends on the careful analysis of the language being tested and the particular course objectives in curriculum (Henning, 1987). Therefore, the greater a test's content validity is, the more likely it is to be an accurate measure of what it is supposed to measure and to find out those areas ignored in language teaching and learning so as to avoid negative washback as much as possible (Hughes, 2003).

It is necessary to consider the content validity in the time of designing and developing a test, evaluation on content validity should not wait until the test is already being tested (Hughes, 2003). However, it is also quite significant to carry out content validation after organizing a test in order to find out shortage of that test. A common way to carry out content validation is to compare it with a statement of what the content ought to be. Such a content statement can be included the teaching syllabus or curriculum, domain specification or test specification (Alderson, Clapham, & Wall, 1995). Hughes (2003) presented that the best way to make sure the test content is to give full test specification, that is describing clearly and precisely the particular language skills and areas that should be included in the test.

A test can achieve good content validity if it contains nothing but the skills and areas that test designers want to be included in a test. However, it is impossible to have everything in the specifications. From the view of statistics, only if test items are random and big enough, they can represent the whole. What the test designers can do is just follow a standard procedure of random selection of test items to make sure that the tasks in the test are representative in that domain (Hughes, 2003).

교보문고 KYOBO Book Centre

2. Major Public English Tests in China

It is believed that all tests, especially public tests used to make important decisions, have great effects on students, teachers, parents, administrators, and test-takers in particular (Messick, 1989, 1996; Paris, Lawton, Turner & Roth, 1991). As a public standardized test, college entrance examination in China, named as the National Matriculation Test (NMT), is the main concern for most parents and the society because the test grade is the only criteria used to decide college admission and college enrollment is considered the most promising way to ensure the students' future employment.

Language tests are no exceptions so that great attention has been put on them. In China, major public English tests designed domestically include the College English Test (CET), the National Matriculation English Test (NMET), the Test for English Majors (TEM), the Graduate School Entrance English Examination (GSEEE), the Public English Testing System (PETS), the Cambridge Business English Certificate (BEC), and the WSK (abbreviation of Chinese phonetic alphabet, which means the national English language proficiency test), used to select professionals for studying and training overseas (Cheng, 2008).

The NMET, as one of the most important parts of the NMT, is the college entrance examination of English. It is one of the largest standardized norm-referenced proficiency EFL (English as a foreign language) tests used to make inferences about candidates' English language ability in college admission decisions (Qi, 2004). Since its formulation in 1950, there have been significant changes in terms of test item types, content, testing methods, and ways of scoring. The fifty-three years' development of the NMET reflect our understanding and adoption of language testing theories during the four periods of language testing, including pre-scientific testing development, psychometric-structural testing, psycholinguistic-sociolinguistic testing or integrative testing, and the last but not the least, communicative testing at its trial stage, also called semi-communicative testing (Lu & Liu, 2008).

In terms of the contents tested, the NMET has also undergone significant changes, from testing reading and translation to testing textual and pragmatic competence. The testing methods have become much broader thanks to the adoption of the multiple-choice questions. As Song and Jin (2003) stated that all these changes have taken place in order to reflect the standards set by the English curriculum of different periods.

교보문고 KYOBO Book Centre

III. METHODS

1. Research Objectives

The objective of this study is to investigate content validity of the language knowledge sections of four representative NMET papers in order to see to what extent they meet the requirements of the NSECS on language knowledge. More specifically, this study addresses the following questions:

1) What are the characteristics of the language knowledge sections of four NMET papers in terms of contents, item types, and the number of test items?

2) What are the differences between four NMET papers in terms of the language knowledge sections? Which one better reflects the NSECS?

3) What are the test item quality and consistence of the language knowledge sections of four NMET papers?

4) What are the factors leading to the poor quality?

2. The NMET Papers for Investigation

Due to the decentralization of the NMT administration by the Ministry of Education in China, most provinces or municipalities have begun to design their own NMET papers by implementing new English curriculums for senior high schools. During the 2012 academic school year, sixteen provinces or municipalities have designed their own versions of the NMET. Including three national versions, NMET version I, II and new English curriculum version, the number of the NMET versions adds up to nineteen each with distinct characteristics in terms of item types, content and testing methods. Among them, the NEC, Beijing, Shanghai and Guangdong versions were selected for analysis due to their great influences in field of the NMET (http://edu.qq.com).

3. Instruments

Since the study is to investigate content validity of the language knowledge sections of the NMET papers according to the NSECS, it is necessary to compare the test content with the specification of the NSECS and test syllabus. According to the new English curriculum, the goal of English teaching is to develop students' comprehensive English language competence through five general objectives including language skills, language knowledge, emotional attitudes, learning strategies, and culture awareness. Therefore, the instruments used in the present study were the overall descriptors of the NSECS and test syllabus for

교보문고 KYOBO Book Centre

level 8[1] students on language knowledge. This study analyzed the content validity of language knowledge sections of four NMET papers administered in 2012 by examining their framework of task characteristics proposed by Bachman and Palmer (1996). The classification of testing the test point validity proposed by Li (1997) was also adopted to analyze the test points in language knowledge section of four NMET papers, including the NMET version of NEC, Beijing, Shanghai and Guangdong. Basic grammar and grammatical use of vocabulary specified in the NSECS for level 8 students and test syllabus of 2012 were discussed in details in section four.

IV. RESULTS AND DISCUSSION

Bachman (1990) claimed a language test as "If we are to develop and use language tests appropriately, we must base them on clear definitions of both the abilities we wish to measure and the means by which we observe and measure these abilities" (p. 81).

To analyze a language test, it is also essential to specify both the language abilities measured and the methods employed to evaluate those abilities so that Bachman (1990) proposed the framework of test method facets, which was defined as a framework of task characteristics by Bachman and Palmer (1996), as shown in Table 1 below.

TABLE 1

Framework of Task Characteristics (Bachman & Palmer, 1996, p. 57-58)

Characteristics of the setting	Physical characteristics; Participant; Time of task	
Characteristics of the test rubric	Instructions	Language (native, target) Channel (aural, visual) Specification of procedures and tasks
	Structure	Number of parts/asks Salience of parts/asks Sequence of parts/asks Relative importance of parts/asks Number of task items per part
	Time allotment	
	Scoring method	Criteria for correctness Procedures for scoring the response Explicitness of criteria and procedures

[1] By following the international practice of dividing the general objectives into different ability levels, the New English Curriculum Standards designs nine levels from Grade 3 in primary school to Grade 3 in senior high school. Level 8 is the required standard for senior high school graduates, while level 9 is an extension level for specialist schools and able students.

교보문고 KYOBO Book Centre

Characteristics of the input and the expected response	Format	Channel (aural, visual) Form (language, non-language, both) Language (native, target, both) Length Type of input (item, prompt) Type of expected response (selected, limited production, extended production) Degree of speededness Vehicle (live, reproduced, both)
	Language of input and expected response	Language characteristics -Organizational characteristics Grammar (vocabulary, syntax, phonology, graphology) Textual (cohesion, rhetorical/conversational organization) -Pragmatic characteristics Functional (ideational, manipulative, heuristic, imaginative) Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)
	Topical characteristics	
Relation between input and response	Reactivity (reciprocal, non-reciprocal, adaptive) Scope of relationship (broad, narrow) Directness of relationship (direct, indirect)	

Content validity analysis of four NMET papers was conducted based on the framework of task characteristics by Bachman and Palmer (1996), however, they also presented that test developers or researchers could make some modifications in the specific characteristics for their own purposes. The four NMET papers investigated in this study are based on the test syllabus and specifications of the NMET under the influence of the NSECS. However, test content, items, types and testing method of each NMET paper are quite different from each other. Thus, only some categories of the framework of task characteristics illustrated in Table 1 were used to analyze the content validity of four NMET papers due to the research needs of this study, including the analysis of characteristics of the setting, characteristics of the test rubric, and some categories in the aspect of characteristics of the expected response.

1. Characteristics of the Setting

The testing environment, physical condition, is similar for all the test takers, and all the test takers are familiar with the testing environment, for they are granted to familiarize themselves with the testing place at the certain time before taking the test. Test takers usually take the NMET in the classrooms, traffic controls are in operation around the area of school campus and automobiles are not allowed to drive through which ensures the necessary quiet environment for the test. For the participants, there are usually 30 test takers and 2 exam supervisors in each classroom. The two exam supervisors administering

교보문고 KYOBO Book Centre

and presiding over the test are supposed to be totally unfamiliar to reduce possible opportunities of cheating. The NMT is held from June 7th to June 9th every year. The NMET is conducted on the afternoon of June 8th, usually from 3 p.m. to 5 p.m., for two hours to make sure the suitable temperature and humidity to the test takers on that day.

2. Characteristics of the Test Rubric

Comprehensive English language competence is the ultimate goal for learners to achieve according to the general objectives of new English curriculum. The NMET should also be designed to reflect learners' comprehensive English language competence.

Test instructions of the NMET versions of NEC, Beijing and Guangdong were presented in the same language, the test takers' native language Chinese, while test instructions of the NMET paper of Shanghai was given with the target language English through a visual channel before each type of test items. Of course in listening comprehension, the aural input was given to the students together with the visual instructions. Unlike the difference on instructions, all test items and selecting options were presented in English on the four NMET papers, except the test content of translation of Shanghai NMET, in which test takers were asked to translate five Chinese sentences into English, providing each sentence with keyword in English at the end. The special test item is writing, in which all instructions and requirements for writing were stated in native language Chinese in order to avoid test takers' misunderstanding and guide them complete writing task well.

Each NMET paper consists of several sections among the following aspects, like listening comprehension or listening & speaking test, language knowledge, reading comprehension, translation, and writing. The specific analysis of the structures of the NMET versions of NEC, Beijing, Shanghai, and Guangdong in terms of total number of questions, test content, parts and number of questions with type marking, total score, and time allotment is demonstrated in the following tables, as shown from Table 2 to Table 5.

The NMET version of NEC is composed of two parts, named paper 1 and paper 2 with the time allotment of 120 minutes. Test items in paper 1, covering the test content of listening comprehension, language knowledge and reading comprehension, are all objective multiple-choice questions marked by italic type to choose the most suitable answer out of four possible options, except test items of passage completion, choosing five possible options out of seven into five blanks underlined respectively. Paper 2 contains subjective test items concerning with test content of writing with two subtests, proof-reading, a kind of error correction in a paragraph and situational writing with given instructions and requirements in Chinese, as shown in Table 2.

교보문고 KYOBO Book Centre

TABLE 2

Structure of the NMET Version of NEC

NMET	Paper NO.	Total number of questions	Section	Parts & number of questions (italic numbers=MCQ)		Total score		Time (min)
NEC	Paper 1	75	Listening	Short conversation	*5*	7.5	150	120
				Long conversation	*15*	22.5		
			Language knowledge	Gra & Voc	*15*	15		
				Cloze test	*20*	30		
			Reading	Reading	*15*	30		
				Passage Completion	*5*	10		
	Paper 2	11	Writing	Proof-reading EC(p)	10	10		
				Situational writing	1	25		

Note: Version NEC=the NMET version of new English curriculum for national use, MCQ=multiple-choice questions, Gra & Voc=grammar and vocabulary, EC (p)=error correction in paragraph

TABLE 3

Structure of Beijing NMET

NMET	Paper NO.	Total number of questions	Section	Parts & number of questions (italic numbers=MCQ)		Total score		Time (min)
Beijing	Paper NMET	75	Listening	Short conversation/dialogue	*5*	7.5	150	120
				Long conversation/dialogue	*10*	15		
				Dictation of BF (w)	5	7.5		
			Language knowledge	Gra & Voc	*15*	15		
				Cloze test	*20*	30		
			Reading	Reading	*15*	30		
				Passage Completion	*5*	10		
		2	Writing	Situation writing	1	20		
				Open writing	1	15		

Note: MCQ=multiple-choice questions, BF=blank filling, Gra & Voc=grammar and vocabulary

Beijing NMET isn't divided into two parts, however, the test content, items and types are quite similar with those of the NMET version of NEC including listening comprehension, language knowledge, reading comprehension and writing with the time allotment of 120 minutes. Unlike the test items in listening comprehension designed with objective multiple-choice questions on the NMET version of NEC, there are test items of dictation with five questions on Beijing NMET, filling in each blank with one word after listening to a conversation. The test items in language knowledge and reading comprehension are the same as those on the NMET version of NEC. However, test takers are required to write two compositions on Beijing NMET. One is situational writing with given instruction in

교보문고 KYOBO Book Centre

Chinese and a set of pictures with four pieces and the other is open writing which requires test takers to write their understanding about a picture to his English friend Jim, as shown in Table 3.

Like the NMET version of NEC, Shanghai NMET also consists of two parts, named paper 1 and paper 2 with the time allotment of 120 minutes. Test items in paper 1, covering the test content of listening comprehension, language knowledge and reading comprehension, are all designed with objective multiple-choice questions except test items of dictation in listening comprehension and reading for questions and answers or statement completion in reading comprehension. Dictation test of Shanghai NMET is divided into two parts, the first four questions were required to fill in the blank with one word, the other four were asked to fill in the blank with no more than three words after listening to a longer conversation. In the test item of reading for questions and answers or statement completion, there are two questions to answer and two statements to complete with the information in the passage by filling in the blank with no more than eight words. The special one is the test content of language knowledge, in which grammar and vocabulary are separately measured, with the type of multiple-choice questions for grammar and filling in the blank with nine optional words out of ten respectively. Unlike cloze test classified into the parts of language knowledge on the versions of NEC, Beijing and Guangdong, it belongs to the part of reading on Shanghai NMET. In the subtest of reading comprehension, there are test items requiring test takers to choose five headings out of six best suitable for five short passages, as shown in Table 4.

TABLE 4

Structure of Shanghai NMET

NMET	Paper NO.	Total number of questions	Section	Parts & number of questions (italic numbers=MCQ)		Total score	Time (min)
Shanghai	Paper 1	84	Listening	Short conversation	*10*	10	150	120
				Passage	*6*	12		
				Dictation of BF (w)	8	8		
			Language knowledge	Grammar	*16*	16		
				Vocabulary of cloze BF	*9*	9		
			Reading	Cloze	*15*	15		
				Reading	*11*	22		
				Reading for choosing heading	*5*	5		
				Reading for Q & A or completing statement	4	8		
	Paper 2	6	Translation	Sentence translation	5	20		
			Writing	Guided writing	1	25		

Note: MCQ=multiple-choice questions, BF=blank filling, Q & A=questions and answers

교보문고 KYOBO Book Centre

Table 4 also shows that the items in paper 2 are all subjective including sentence translation and guided writing. The most impressive one is sentence translation, in which test takers are asked to translate five Chinese sentences into English, given one key word in English at the end of each sentence. The same as situational writing, guided writing asks test takers to write a composition with given instructions and requirements in Chinese along with two pieces of pictures.

Table 5 shows that test takers in Guangdong takes the NMET test twice, first for listening and speaking test in the mid of March, second for pen and paper test on the afternoon of June 8th. Listening and speaking test is called computer-based English listening and speaking test. It is the first attempt for designing the NMET incorporating listening and speaking test. That is, only Guangdong NMET includes the speaking test. As the name shows, it requires test takers to complete the test on computer for about 30 minutes including reading after a speaker after watching a video clip named reading aloud, asking three questions and answering five questions called role play, and retelling what they heard from a monologue defined as retelling. Though the total score of listening and speaking test on the paper is 60, it will be recalculated as 15, accounting for 10% of total score 150.

TABLE 5

Structure of Guangdong NMET

NMET	Paper NO.	Total number of questions	Section	Parts & number of questions (italic numbers=MCQ)		Total score		Time (min)
Guangdong	Paper L & S	10	Listening & Speaking	Reading aloud	1	5	15	30
				Role play	8	4		
				Retelling	1	6		
	Paper NMET	50	Language knowledge	Cloze test	*15*	30	135	120
				Cloze BF for Gra & Voc	10	15		
			Reading	Reading	*20*	40		
				Information matching	*5*	10		
		2	Writing	Situational writing	1	15		
				Task writing	1	25		

Note: MCQ=multiple-choice questions, Gra & Voc=grammar and vocabulary

The second part, pen and paper test covers the test content of language knowledge, reading comprehension, and writing with the time allotment of 120 minutes. The special one is that there is no test items with the type of multiple-choice questions to measure learners grammar and vocabulary ability, while this ability is checked according to cloze

교보문고 KYOBO Book Centre

test with the type of blank filling, some required to fill in the suitable word with right tense or voice of given words, some asked to insert new words considering the context.

Unlike other reading comprehension test, there is information matching test in the subtest of reading comprehension on Guangdong NMET, in which test takers were required to match five optional information out of six with specific letters written to an editor. There are two tasks in the test content of writing. One is situational writing with given instructions and requirements in Chinese and the other is task writing for summarizing the main idea of a passage and writing a composition of test takers' opinions based on the given instructions in Chinese, as shown in Table 5.

Though the total time allotment is all given to the four NMET versions, the time required for each test item is not provided. Except the listening and speaking test of Guangdong NMET, in which test takers were required to complete the test by following the requirements of speaker, test takers should allocate the time for each subtest and test items belong to it themselves. Actually, it would affect the test performance of some test takers who may spend too much time on certain questions, while leaving insufficient time to complete the other questions.

3. Characteristics of the Expected Response

Due to the research objective of this study, the main focus is put on the characteristics of the expected response in terms of response types and subskills covered on the NMET versions of NEC, Beijing, Shanghai and Guangdong to discuss content validity of the language knowledge sections of four NMET papers from the perspective of the requirements of the NSECS.

According to Song and Jin (2003), as one of the important components of comprehensive English language competence, language knowledge is classified into the following five parts, phonetics, grammar, vocabulary, functions and topics. It is important to be clear about the differences between learning objectives (curriculum) and learning outcomes (assessment). That is, not all objectives can be covered or measured in a single test. It is always the case of selecting the key learning objectives in order to measure their learning outcomes, especially like a large scale and standardized test, the NMET. Thus, in the test items of language knowledge, basic grammar and vocabulary are the main concerns and measured in the NMET all the time. However, the test item types are various from each other, which were discussed in the section of the characteristics of test rubric of four NMET versions above.

The general objectives required by new English curriculum for level 8 students on language knowledge in terms of grammar are as follows (Song & Jin, 2003):

1) Be able to further master how to describe time, location and position.

교보문고 KYOBO Book Centre

2) Be able to further understand and master expressions on how to compare people, objectives and things.

3) Be able to describe and express opinions, attitudes, and emotions with proper language forms.

4) Be able to learn and master the basic knowledge of discourse, and be able to organize information effectively according to particular objectives.

While the general objectives required by new English curriculum for level 8 students on language knowledge in terms of vocabulary are as follows (Song & Jin, 2003):

1) Be able to use vocabulary to understand and express different functions, intentions and attitudes, etc.

2) Be able to use vocabulary to describe relatively complicated things, behaviors and characteristics, and explain the principles, etc.

3) Be able to learn and use 3000 words and 400-500 expressions and collocations.

In analyzing types of responses in language knowledge sections of four NMET papers, Bachman and Palmer's (1996) classification of expected response was adopted. Based on Bachman's (1990) classification of response type into selected response and constructed response, Bachman and Palmer (1996) pointed out that the expected response consists of selected response, limited production response and extended production response by considering the knowledge of cohesion and rhetorical organization in utterances, which should be separated from the single sentence responding form. The selected response is designed by multiple-choice questions, which requires test takers to choose the best one suitable to the context out of several possible options. The limited production response refers to those items needs to fill in words like blank filling, sentence completion with a word, phrase, or sentence. However, if the response is composed of more than one sentence, it is called the extended production response. Among three types of response, extended production response requires the highest ability of language competence, while selected response requires the least in general.

The distribution of three types of responses on the NMET versions of NEC, Beijing, Shanghai and Guangdong in terms of language knowledge test is provided, as shown in Table 6.

TABLE 6

Types of Responses in Language Knowledge Sections of Four NMET Papers

The NMET versions	Selected production	Limited production	Extended production
NEC	35		
Beijing	35		
Shanghai	25		
Guangdong	15 (Cloze with type of MCQ)	10 (Cloze BF for Gra & Voc)	

Note: Version NEC=the NMET version of new English curriculum for national use, MCQ=multiple-choice questions, BF=blank filling, Gra & Voc=grammar and vocabulary

교보문고 KYOBO Book Centre

Seen from Table 6, test items in language knowledge are all designed with the type of multiple-choice questions except the test items of cloze BF which require test takers of Guangdong to fill in the words with right tense and voice best suitable to the context with given words or just insert new words according to the context. Thus, the response type of test items of language knowledge sections of the NMET version of NEC, Beijing and Shanghai are all selected types, while two types of responses can be found in the language knowledge test of Guangdong NMET, selected and limited production types. No response type of extended productions is found on the four NMET papers. Though selected response types designed by multiple-choice questions can make sure the high reliability of the test, saving time, money and labor, the disadvantages of multiple-choice items themselves shouldn't be neglected. Those items are lack of interaction with the test and test takers, also it is inevitable to avoid the chances of guessing so that like Guangdong NMET, limited production response types should be increased to some extent, so do the extended production response types.

In the 2012 test syllabus about the NMET, there are two requirements for test takers to complete the language knowledge task in terms of the following aspects:

1) Be able to master and use phonetics, vocabulary, basic knowledge of grammar, and already learned functions and topics

2) Be able to learn and use about 3500 words

In the test content of language knowledge sections, two main types of test items are employed on the NMET versions of NEC, Beijing, Shanghai and Guangdong, multiple-choice items for grammar or vocabulary, or both, cloze test for grammar or vocabulary, or both. However, test items with the same type are quite different from each other among those four NMET papers, as shown in Table 7.

TABLE 7

Test Items in the Part of Language Knowledge Sections of Four NMET Papers

NMET	Items	Types	Description
NEC	Multiple-choice	MCQ	Grammar and vocabulary
	Cloze test	MCQ	Grammar and vocabulary
Beijing	Multiple-choice	MCQ	Grammar and vocabulary
	Cloze test	MCQ	Grammar and vocabulary
Shanghai	Multiple-choice	MCQ	Grammar and vocabulary
	Cloze test	MCQ	Vocabulary (Choose nine suitable words out of ten possible options into nine blanks)
Guangdong	Cloze test	MCQ	Grammar and vocabulary
	Cloze test	BF	Grammar

Note: MCQ=multiple-choice questions

교보문고 KYOBO Book Centre

Table 7 shows that items of multiple-choice and cloze test with the type of multiple-choice questions are employed to measure test takers' language knowledge of grammar and vocabulary. Items of multiple-choice with the type of multiple-choice questions are also accepted by Shanghai NMET to measure test takers' ability of grammar and vocabulary. While another test items of cloze test, choosing nine suitable words out of ten possible options into nine blanks, are employed to measure the ability of learning and using vocabulary as well. No items of multiple-choice are found on the NMET paper of Guangdong since Guangdong NMET adopted two cloze tests; one is used to measure test takers' abilities of grammar and vocabulary, and the other is employed to measure test takers' abilities of grammar through filling in the blanks with the right tense and voice with given words, or just inserting new words according to the context.

Multiple-choice question is one of the most popular item types widely used in the field of educational assessment, which requires test takers to choose the best suitable answer out of several possible options. Good multiple-choice questions are said to be valid because they can measure a variety of language abilities. However, most multiple-choice questions are designed to measure only one discrete language point, especially those multiple-choice items used to measure test takers' learning and using abilities of vocabulary and basic knowledge of grammar. Multiple-choice questions in cloze test can be designed quite different, putting the main focus on measuring test takers' integrative language abilities in a particular context.

Cloze test is an item type in which words are systematically removed from a text and replaced with blanks, and the test takers' task is to fill in the best suitable words according to the context (Read, 2000). The test designers need to find out a suitable passage which agrees with the test takers' language abilities and delete or remove every n-th word considering the facts that the test is used not only to measure one language point a time, but also to measure test takers' integrative language abilities. Test takers should employ all his or her language knowledge covering grammar, vocabulary, functions and topics, skills of reading, learning strategies, culture awareness, and testing strategies as well to complete the task of cloze test. Cloze test measures the integrative ability of test takers. It can be said to have a high content validity since it covers more than one ability it intends to measure.

Due to the various types of test items of language knowledge sections on four NMET versions of NEC, Beijing, Shanghai, and Guangdong, this paper just analyzed the test points in the part of language knowledge sections as a whole without dividing the items into multiple-choice and cloze test. Test points of grammar in language knowledge can be classified into two parts in terms of basic knowledge grammar and grammatical use of vocabulary according to the test specification of the NMET test syllabus of 2012.

In addition, the classification of testing the test point validity presented by Li (1997) was employed to discuss the test points in the part of language knowledge in the following

교보문고 KYOBO Book Centre

terms of word level, phrase level, sentence level, and discourse level. As word level is the lowest level, no context is needed when deciding the choice. Phrase level means those words, usually in the form of collocations, are chosen by information before or after the blank. Sentence level refers to those words which are decided within a sentence. While in discourse level, words are decided not only by a sentence, but also by the whole passage. Li (1997, p. 309) further commented that discourse level test point is definitely more capable than the word level test point in measuring the real language performance of test takers. Generally, the higher level of test point represents the higher level of validity, that is, the high level test point can measure low level abilities simultaneously, but not vice versa. Based on the above classification, the test points in the part of language knowledge sections of four NMET papers are shown in Table 8.

The 2012 NMET test syllabus provided 24 test points of grammar, including 8 different classes of words in terms of notional and functional words, tense, passive voice, nonfinite verb, word-formation, sentence types, sentence constituents, basic patterns of simple sentence, subject-verb agreement, compound-complex sentences, complex sentence, indirect speech, ellipsis, inversion, emphatic patterns and subjunctive mood. The test point of communicative context is included in Table 8 in this study in order to see whether those multiple-choice test items are designed in real life situation, though it is not specified in the test points of grammar in test syllabus.

TABLE 8

Test Points of Language Knowledge Covered on the Four NMET Papers

Classification	Test points	NEC 35 questions	Beijing 35 questions	Shanghai 25 questions	Guangdong 25 questions
Vocabulary	Noun	8 (*37*, *42*, *44*, *46*, *50*, *53*, *54*, *55*)	9 (*36*, *39*, *40*, *46*, *48*, *51*, *52*, *54*, *55*)	9 (*42*, *43*, *45*, *47*, *50*, *51*, *60*, *62*, *63*)	3 (*3*, *4*, *7*)
	Pronoun	1 (31)		1 (31)	1 (*25*)
	Preposition		2 (34, *50*)	2 (25, *56*)	1 (*23*)
	Conjunction	4 (25, 34, *41*, *45*)	2 (21, *43*)		1 (*20*)
	Adjective	8 (26, 29, *36*, *39*, *43*, *48*, *49*, *52*)	3 (28, *45*, *53*)	7 (*41*, *46*, *49*, *54*, *58*, *61*, *64*)	7 (*1*, *5*, *9*, *11*, *14*, *18*, *19*)
	Adverb	1 (*40*)	1 (*38*)		4 (*2*, *6*, *8*, *10*)
	Article	1 (22)			1 (*19*)
	Verb (verb phrase, modal verb)	6 (27, 30, 35, *38*, *47*, *51*)	8 (31, 33, *37*, *41*, *42*, *44*, *47*, *49*)	8 (28, *44*, *48*, *52*, *53*, *55*, *57*, *59*)	3 (*12*, *13*, *15*)
Grammar	Tense	2 (23, 33)	4 (22, 25, 29, 30)	2 (26, 30)	

교보문고 KYOBO Book Centre

	Passive voice		1 (29)	1 (26)	
	Nonfinite verb	2 (28, 32)	2 (23, 27)	4 (27, 33, 36, 39)	1 (*16*)
	Subject-verb agreement		1 (35)		
	Clause		2 (24, 26)	5 (32, 34, 35, 38, 40)	2 (*22*, *24*)
	Inversion			1 (29)	
	Emphasis	1 (24)		1 (37)	
	Subjunctive mood				1 (*17*)
	Communicative context	1 (21)	2 (21, 29)	3 (30, 37, 38)	
Gradation	Word level	0	0	0	0
	Phrase level	1 (29)	3 (*41*, *47*, *54*)	3 (27, 33, *49*)	2 (*7*, *19*)
	Sentence level	19 (22, 23, 24, 25, 26, 28, 30, 31, 32, 33, 34, 35, *37*, *39*, *40*, *44*, *52*, *54*, *55*)	18 (22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, *37*, *48*, *49*, *50*, *51*)	13 (25, 28, 29, 31, 32, 34, 35, 36, 39, 40, *47*, *53*, *62*)	11 (*3*, *4*, *13*, *14*, *16*, *17*, *18*, *21*, *22*, *23*, *24*)
	Discourse level	15 (21, 29, *36*, *38*, *41*, *42*, *43*, *45*, *46*, *47*, *48*, *49*, *50*, *51*, *53*)	14 (21, 29, *36*, *38*, *39*, *40*, *42*, *43*, *44*, *45*, *46*, *52*, *53*, *55*)	24 (26, 30, 37, 38, *41*, *42*, *43*, *44*, *45*, *46*, *48*, *50*, *51*, *52*, *54*, *55*, *56*, *57*, *58*, *59*, *60*, *61*, *63*, *64*)	12 (*1*, *2*, *5*, *6*, *8*, *9*, *10*, *11*, *12*, *15*, *20*, *25*)

Note: 5 (16, 17)='5' before bracket refers to the total number of question items, and '(16, 17)' refers to the sequence of question items, items with italic type=items of cloze test, '29'=items measured more than one test point

With careful analysis of the test questions, this paper finds out the following characteristics:

1) Among the NMET versions of NEC, Beijing and Shanghai, the test points of multiple-choice test items put emphasis both on grammatical use of vocabulary and basic knowledge of grammar. However, the ratio is different among the three NMET papers. The largest proportion given to measure test takers' basic knowledge of grammar is Shanghai NMET, in which 13 questions are measured out of 16 ones. While the least one is the NMET version of NEC, 5 questions are measured out of 15 ones.

2) All cloze test items on four NMET papers mainly focus on the grammatical use of vocabulary, especially in terms of notional words, like noun, adjective, and verb concerned with the meaning discrimination and logical judgment according to the context. No questions in cloze test are found to measure basic knowledge of grammar on the NMET versions of NEC, Beijing and Shanghai. Only 4 questions are found on Guangdong NMET paper to measure the basic knowledge of grammar in cloze test of blank filling due to the

교보문고 KYOBO Book Centre

special design of test items in Guangdong NMET, employing two different types of cloze tests to measure test takers' language knowledge abilities.

3) In multiple-choice items, two questions are found to measure more than one test point of test takers' basic knowledge of grammar. They are question sequence NO. 29 on the NMET paper of Beijing and question sequence NO. 26 on the NMET paper of Shanghai, which are employed to measure tense and passive voice both at a time. Though test takers' integrative knowledge of grammar can be measured in a multiple-choice question, but the number of such questions is too small.

4) Unlike traditional multiple-choice questions measuring test takers' abilities of language knowledge in the unreal context, it is surprising to find that 6 questions in total are designed in a real life communicative context, 1 in version of NEC, 2 in Beijing, 3 in Shanghai NMET papers, respectively. Especially, the multiple-choice question sequence NO. 21 on the NMET version of NEC is not used to measure a certain test point of grammatical use of vocabulary and basic knowledge of grammar. On the contrary, it is employed to measure test takers' communicative competence which requires test takers to choose a suitable answer out of four possible options according to the context.

5) From the results of test point gradation, no word level test points are found. Among four kinds of test point levels, sentence level and discourse level test points are the major ones, and only a small number of phrase level test points are accepted. However, the distribution of sentence and discourse level test points in different types of test items is quite different.

6) The test points in multiple-choice items on the NMET version of NEC, Beijing and Shanghai and cloze with the type of blank filling on the NMET paper of Guangdong are mostly designed in sentence level. However, the trend of designing test points into discourse level can also be found in those test items above. 10 questions in total are found designed into discourse level test points, 2 in version of NEC, 2 in Beijing, 4 in Shanghai and 2 in Guangdong NMET papers, respectively. It is worth mentioning that discourse level test points have to require test takers to consider the fact of context in mind when giving or choosing the right answers. That is, the basic knowledge of grammar or grammatical use of vocabulary is incorporated into context, which involves test takers' high level ability of language knowledge.

7) The test point level of cloze tests with the type of multiple-choice questions on the NMET versions of NEC, Beijing, Shanghai and Guangdong mainly focus on the discourse level. Test takers have to make full use of his integrative ability to complete the task, including language knowledge, reading skills, strategies in learning and testing, etc. Therefore, from the view of test point levels, cloze test is closely consistent with the requirements of the new English curriculum to explore test takers' comprehensive English language competence.

교보문고 KYOBO Book Centre

Seen from Table 8, among 24 test points of grammar stated in the test specification of the 2012 NMET test syllabus, 17 test points are employed to measure test takers' ability of language knowledge in terms of basic knowledge of grammar and grammatical use of vocabulary on the NMET versions of NEC, Beijing, Shanghai, and Guangdong. Comparing the test points of grammar and vocabulary on four NMET papers with test points in the 2012 test syllabus, all of the test points in the part of language knowledge sections of four NMET papers are included in the 2012 test syllabus. Considering test syllabus and specification designed based on the requirements of the NSECS, it can be said that test content of grammar and vocabulary in the part of language knowledge sections has achieved agreement with subskills required by the NSECS.

V. CONCLUSION

With careful analysis and discussion of the language knowledge sections of four representative NMET versions of NEC, Beijing, Shanghai and Guangdong administered in 2012, this study has shown that all of the language knowledge sections of four NMET papers had relatively high content validity with its design corresponding to the requirements of the NSECS and the 2012 test syllabus in terms of basic grammar and grammatical use of vocabulary. Test items are mainly designed in relatively difficult sentence and discourse level. Even the test points in multiple-choice items have shown the trend to design the test points into discourse level. Among 24 test points of grammar stated in the test specification of the 2012 NMET test syllabus, 17 test points are employed to measure test takers' ability of language knowledge in terms of basic knowledge of grammar and grammatical use of vocabulary on four NMET papers. Though the formats of four NMET papers were slightly different, all NMET designs laid stress on the notion of language use, and they not only examined test takers' basic knowledge of grammar and grammatical use of vocabulary, but also evaluated their ability to use them in a given task and particular setting, that is, the comprehensive language competence.

Some deficiencies are found which might affect content validity of the NMET. In terms of test content of language knowledge sections, due to the test items are all designed with the multiple-choice questions, the expected responses are all limited to selected production types, except test items of cloze test with type of blank filling which requires test takers of Guangdong to fill in the right tense and voice best suitable to the context with given words or just insert new words with the response types of limited productions. No extended production responses are found on four NMET papers. Since the important factors of test formats and test methods also affect test interpretation (Bachman, 1990; Bachman &

교보문고 KYOBO Book Centre

Palmer, 1996; Shin, 2012), an attempt to standardize the test items of test format and test methods is required to enhance the validity of the test as well as the reliability of the test.

Thus, the proportion of subjective test items, such as limited production and extended production responses, should be increased to test high-order level of language knowledge as a communicative testing to evaluate test takers' comprehensive English language competence. Even in designing test items with the type of multiple-choice questions, innovations can be made to diversify the types of test items, like cloze test with the type of blank filling in Guangdong NMET.

REFERENCES

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. London: Longman.

American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, D. C.: American Psychological Association.

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

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and development useful language tests*. Oxford: Oxford University Press.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices.* New York: Pearson Education.

Cheng, Liying. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15-37.

Dong, Manxia. (2011). A diachronic study of the content validity of reading comprehension of Chongqing NMET from 2004 to 2009. *Educational Measurement and Evaluation, 2011*(2), 22, 52-56.

Gu, Xiangdong, & Gao Xiaoying. (2007). A synchronic study of the writing items of the 2007 NMET test papers. China Examinations *(Examination Research Edition), 2007*(12), 28-33, 36.

Gu, Xiangdong, & Lei, Xuemei. (2009). A study of the listening items of the 2008 NMET test papers. *Educational Measurement and Evaluation, 2009*(11), 48-52.

Gu, Xiangdong, Yang, Peirong, & Lei, Xuemei. (2009). A study of the quality of 19 sets of writing items of 2009 NMET test papers. *Educational Measurement and Evaluation, 2010*(12), 47-50.

Henning, G. (1987). *A guide to language testing: Development, evaluation and research.* Cambridge, MA: Newbury House.

교보문고 KYOBO Book Centre

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Jeon, Byoung-Man. (2004). A comparative study of English test items of college entrance examinations in Korea, China and Japan. *English Language & Literature Teaching, 10*(2), 113-132.

Li, Xiaoju. (1997). *Science and art of language testing*. Changsha: Hunan Education Press.

Lu, Qiaoling, & Liu, Sujun. (2008). A study of the development of English language testing in China. *Journal of Guangxi Normal University: Philosophy and Social Sciences Edition, 44*(2), 94-99.

Madaus, G. F. (1988). The influence of testing on the curriculum. In L. Travers (Eds.), *Critical issues in curriculum: Eighty-seventh yearbook of the national society for the study of education* (pp. 83-121). Chicago, IL: Chicago University Press.

McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Messick, S. (1989). Validity. In R. Linn (Eds.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Oller, J. W. (1979). *Language tests at school*. London: Longman.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*(5), 12-20.

Qi, Luxia. (2004). *The intended washback effect of the National Matriculation English Test in China: Intentions and reality*. Beijing: Foreign Language Teaching and Research Press.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Shin, You-Sun. (2012). A validity study on the vocabulary grade levels test for Korean elementary students. *English Language & Literature Teaching, 18*(2), 125-147.

Song, Guiyue, & Jin, Ying. (2003). *A teacher's manual of the new English curriculum standards for senior high schools in full-time compulsory education* (2nd ed.). Wuhan: Central China Normal University Press.

Weir, C. J. (1990). *Communicative language testing*. New York: Prentice Hall.

http://edu.qq.com/zt2012/2012gkst/index.htm.

교보문고 KYOBO Book Centre

Examples in: English
Applicable Language: English
Applicable Levels: Secondary

Ming-Hao, Jin
Department of English
Yanbian University
977 Gongyuan Road, Yanji 133002, Jilin Province, P. R. China,
Tel: +86-433-273-2412/C. P. :+86-156-9958-8978
Email: jinminghao@ybu.edu.cn, jinminghao25@jbnu.ac.kr

Byoung-Man, Jeon
Department of English
Chonbuk National University
567 Baekje-daero, deokjin-gu, Jeonju 561-756, Jeollabuk-do, Korea
Tel: 063-270-3210
Email: bmjeon@jbnu.ac.kr

Received in October 15th, 2013
Reviewed in November 26th, 2013
Revised version received in December 5th, 2013

교보문고 KYOBO Book Centre