

## CONTENT VALIDITY STUDY ON READING COMPREHENSION TESTS OF N MET

Dong Manxia  
Yangtze Normal University

### Abstract

Based on the framework of task characteristics proposed by Bachman and Palmer (1999), this paper attempts to explore the characteristics and content validity of reading comprehension tests of N MET. By comparing and analyzing the data carefully, the author concludes that reading comprehension parts of N MET have well met the requirements in the two syllabi and have reached a high content validity. However, the author detects some weaknesses and puts forward some suggestions with the purposes for further improving and developing the future reading tests.

### Key words

N MET ; reading comprehension tests ; content validity

### 1 . Context of the Study

Currently, the National Unified Enrollment Examinations ( N U E E ) is the largest scale test in China. This test reaches a population of roughly seven million participants each year. As an important component of the N U E E, the National Matriculation English Test ( N M E T ) is now well established as a large scale standardized test and widely recognized in China due to its scientific approach, consistent marking, rigorous administration and comparable scores. N M E T, as a high stakes test, is used as an agent to change classroom teaching ( Qi 2005 : 42 ). Moreover, it is commonly agreed that high stakes tests produce a domino effect and the test results are used to make important decisions that immediately or directly affect the test takers and other stakeholders. The results from these tests directly affect the teaching and learning of students in the classroom. Therefore, in the development of a high stakes test, which may significantly affect the lives of those who take it, there is an obligation to carry out a full validation exercise before the test becomes operational ( Hughes 1989 : 33 ). At present, there are many validation studies conducted of some large scale tests such as T O F E L, I E L T S, T E M and C E T. Specifically, Zou Shen (1997) conducts a validation study of the test for English major. Yang Huizhong and Weir (1998) conduct a validation study of the National College English Test. The two studies above show that the researchers make a start and a great contribution to the validation study in China.

However, as a largest scale test in China, there are fewer studies of validation concerning the N M E T. Although Xu Zhisuo and Zhang Lifang (2004) conduct a predicative validity of N M E T, it just concerns the test results of N M E T, not involving the N M E T itself. Also, great time is spent evaluating the N M E T every year, but most of the papers / tests are scored with intuitive judgment and some personal opinions of senior English teachers for N M E T; lacking of theoretical basis. So it is quite necessary to conduct a validation study of N M E T. Due to the limitation of the time and energy and the significant role which reading plays in scoring the N M E T, the present study focuses on the content validity of reading comprehension tests of N M E T.

## 2 . Theoretical Concerns of Content Validity

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 ( Henning 2001 : 94 ). Bachman (1990 ) demonstrates that the test covers a given area of content or ability is therefore an essential part of validation . He points out that content validation includes two aspects : content relevance and content coverage . Content validity is an important issue for both language testing and language teaching . Hughes (1989 27 ) states that “the greater a test's content validity is ,the more likely it is to accurately measure what it is supposed to measure ”. A common way to analyze the content validity of a test is to compare it with a statement of what the content ought to be . Such a content statement may be the test specifications ,or it may be a formal teaching syllabus or curriculum ,or it may be a domain specification ( Alderson ,Clapham & Wall 2000 ).

## 3 . Research Methodology

### 3 .1 Assessment Framework

The present study is primarily conducted on the basis of Task Characteristics proposed by Bachman and Palmer (1999 ). However ,when we try to apply it to our reading assessment ,inadequacy has been detected . Moreover ,Bachman and Palmer (1999 ) themselves suggest that it is not their purpose to prescribe any particular types of test tasks or combination of task characteristics . Therefore ,taking account of the nature of reading and characteristics of reading tests ,evaluation can be conducted from the “ Characteristics of the Input ” and “ Characteristics of the Expected Response ”. “ Input ” refers to the materials that the testees are expected to “ process ” and “ respond to ”. In the present study ,input is interpreted as the passages that the readers encounter in the reading tests . It involves length ,reading speed ,new words ,readabilities ,topics and genres . “ Expected response ” suggests ,by its name ,the “ physical response ” the test attempts to elicit by the instruction and the input . The analysis of expected response in this research is conducted from response types and reading skills . However ,other factors and subfactors listed in the original framework have been removed due to their irrelevance pertaining to the present research ,or the difficulty of measurement that they pose on the present research .

### 3 .2 Research Instruments

The research instruments include a teaching syllabus (2000 ) and testing syllabus which are relatively general official statement guiding teaching and testing . NMET is designed on the basis of the teaching syllabus and testing syllabus ,which are two separate syllabi and coexist in senior middle schools in China .

### 3 .3 Data Collection

By conducting thorough and careful examination of reading comprehension parts of NMET during the past five years ,both qualitative and quantitative data have been collected . In collecting the objective data such as the length ,readability and reading speed ,computers have also been adopted in the present study . Whereas ,when collecting some subjective data such as topic ,genre and reading skill ,the author invites three English postgraduates to make a consistent analysis to ensure the data is more objective .

### 3 .4 Data Analysis

By comparing the data of the survey with the requirements of the two syllabi and referring to some relevant research findings at home and abroad ,the author makes specific analysis and evaluation from the characteristics of “ input ” and “ response ”.

## 4 . Analysis and Evaluation

### 4 .1 Characteristics of Input

Bachman (1991 125 ) indicates that “input consists of the information contained in a given test tasks , to which the test taker is expected to respond ”. In the present paper ,the characteristics of input will be described as length ,speed ,new words ,readabilities ,topics and genres .

1 ) Length

A problem that all reading test developers face is how long the texts should be on which they base their tests . Text length is a surprisingly under researched area ( Alderson 2000 :108 ). It is assumed that the length of a passage is a reference to the quantity of words used in that passage . Careful counting is shown in the table below .

Table 4 .1 Length of reading passages from 2003 to 2007

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	03	A
P	1340	1202	1235	1261	1328	1258	1212	1341	1378	1205	1311	1279
Q	571	625	565	573	672	682	655	573	584	651	604	614
T L	1911	1827	1800	1834	2000	1940	1867	1914	1962	1856	1915	1893

P :Passage ;Q :Question ;TL :Total Length ;A :average

Testing syllabus stipulates that the total length of the five passages should be above 1000 words .

Generally speaking ,an obvious feature observed from the table above is that the average words for reading passages from 2003 to 2007 is 1279 ,the minimum words is 1202 ,while the maximum is 1378 .It well ensures the total length of reading tests over 1000 words . However ,as far as reading passage is concerned ,Alderson ( 2000 :256 ) suggests that text chosen is usually between 150 and 350 words in length . Moreover ,Liu Runqing and Han Baocheng ( 2004 ) also state that a passage within 200 500 words is proper for a reading test . In the study ,the writer finds out that all reading passages are confined to a certain range from 167 to 364 words ,not too short or too long ,which also well correspond to the proposals of reading passage length . Therefore ,based on the relevant research findings and the requirements of the teaching syllabus ,conclusion can be drawn that the eleven papers all demonstrate good control of the input capacity ,and they all answer well for the testing syllabus . In other words ,the adequate input capacity is ensured for testees in the reading tests within the time limit set by the official requirements .

However ,it should be noted that there are some weaknesses in the layout of the reading passages . Generally ,reading passages should try to be arranged according to the principle from short to long or “from easy to difficult ” ,but some papers do not abide by this logic ,such as ,in Paper 2004 ( 2 ) ,the two longest passages of five with 311 words and 319 words are located at the beginning of the reading test ,which probably cause the greater pressure on the test takers mind and exert some negative impact on the performance in the reading tests . Moreover ,the longest passage ( Paper 2006 .2B ) and the shortest passage ( Paper 2006 .2C ) are immediately adjacent to each other . Therefore ,we have reason to suggest that a more balanced layout be set out to make sure that test takers won t find it too abrupt to process the reading passages .

2 ) Speed

As for reading speed of NMET ,so far ,there are two calculations of reading speed . Yang & Weir (1998 ) indicate that the designer of test papers ,taking into account of the fact that it takes some time to read the questions items and think about the answers to the questions ,makes a proportion of one to three -quarters of reading the passages to responding to the questions . The testing syllabus stipulates that the total reading input amount of each paper is over 1000 words ,which are required to finish within thirty -five minutes . That is ,within thirty five minutes it actually takes twenty minutes to read the passages and fifteen minutes to respond to the questions . Therefore ,the writer first calculates the reading speed according to the proportion of twenty to fifteen minutes . In addition ,the writer also adopts the usual way to calculate the reading speed ,or the total input amount is divided by reading time . The two calculations are shown in the table below :

Table 4 .2 Comparison of two reading speed from 2003 to 2007

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	03	A
wpm1	67 .0	60 .1	61 .8	63 .1	66 .4	62 .9	60 .6	67 .1	68 .9	60 .3	65 .6	64 .0
wpm2	54 .5	52 .2	51 .4	52 .4	57 .1	55 .4	53 .3	54 .7	56 .1	53 .0	54 .7	54 .1

The teaching syllabus (2000) describes that there are two grades of reading speed for senior middle school graduates. The first grade or low level requires 50-70 wpm which is for the students who have finished English study in the senior middle school. The second grade or relative high level demands 70-80 wpm in daily English reading. Through the first calculation, generally, the average reading speed for each year respectively is approximate 64.0, the minimum speed is 60.3 wpm and the maximum is 68.9 wpm and it shows great consistency from 2003 to 2007. Given the testing situation and the requirements of the two syllabi, the reading speed of 60-70 wpm is acceptable. Generally speaking, reading speed for every paper almost meets the requirement of the teaching syllabus.

However, it is noted that there are comparatively a lot discrepancies among the different papers. In Paper 2004(4) the minimum speed is 60.3 wpm and the maximum is 68.9 wpm in Paper 2004 (3). The disparity is 8.6 wpm between the two papers. That means, within the same reading time limits, Paper 2004 (3) may be more demanding for the testees than the other ones just in terms of reading speed.

Apart from the first calculation, the author also utilizes the second usual calculation method. According to Table 4.2, it can be discerned that reading speed fluctuates from 50 wpm to 60 wpm, which also meets the requirement of the first level (50-70 wpm) stipulated by the teaching syllabus.

By comparing the two reading speeds above, the writer finds out that they just reach the low requirement of the first grade (50-70 wpm), not reach the second grade (70-80 wpm). NMET, as a power and speed test, mainly aims at selecting the qualified candidates for university. Therefore, reading speed should tend to reach the second grade. Based on the analysis, a suggestion may be made that reading speed of NMET needs to be slightly raised. It is because that raising properly the reading speed of NMET can embody the requirements of the teaching syllabus better and help the students to improve the reading abilities.

### 3) New Words

Alderson (2000) indicates that vocabulary plays a very important role in reading tests and it is important to text comprehension and to test performance. Theoretically, the input passage should contain the words and syntactic structures that are included in the teaching syllabus. However, we do find some vocabulary exceeding the two syllabi. A careful counting of the new words is conducted as follows:

Table 4.3 New words and vocabulary beyond the teaching syllabus

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	03
NW %	1.49	1.58	1.05	1.19	1.96	1.51	1.65	2.16	0.87	1.02	1.30
VTS %	3.14	2.98	2.75	2.38	3.39	3.58	3.96	3.06	3.19	2.24	3.89
C & D %	88.6	85.4	94.1	73.3	77.7	80.0	80.8	85.0	84.1	66.6	86.2

VTS :Vocabulary beyond the Teaching Syllabus ;C & D :Compound and Derivative words ;NW :New Words

Nuttle (1982) argues that for intensive reading, it may be acceptable to have "quite a lot of" new words. This "a lot", according to her, is "two or three percent". Furthermore, the teaching and testing syllabi stipulate clearly that the new words of each reading passage should be below 3%.

Through the analysis of the table above, it can be revealed that, eleven papers all have a reasonable amount of new words, which are far below 3% in the vicinity of 0.87 to 1.96. Moreover, as far as every

passage is concerned ,the new words of most reading passages or 95 .6 % are below 3 % . Therefore ,a conclusion can be drawn that most papers or reading passages strike a comparative balance of the new words distribution and have well answered the requirements of the new words below 3 %in the two syllabi .

However ,if a new word is defined as any word that is not included in the teaching syllabus ,the proportion of new words of reading passages exceeds 3 % . Through the careful analysis of the words beyond the teaching syllabus ,the writer finds out a factthat most new words accounting for over 80 %are compound and derivative words . The author just takes Paper 2006(1 ) as an example and presents these words in order to make readers gain a clear understanding of it .In Paper 2006(1 ) ,there are thirty four words beyond the teaching syllabus ,and among of them there are thirty two compound and derivative words which are italicized . They are presented as follows :

actually , misunderstanding , normally , yellowish , especially , meaningful , showroom , pickup , payment , arrival , reasonable , grammar based , self defeating , activity , timetable , halfway , re-think , actually , watercolor , rediscovered , enjoyable , regretful , untold , attractively , coastal , highway , guided , personalized , wildlife , higher up , overview , skybox , ultimate , bay

Therefore ,based on the findings above ,the writer can make some suggestions as follows :

First ,during the reading test ,the testees should attempt to infer the meaning of the new words from the affixes or roots that have been required in the teaching syllabus .Second ,regular teaching should lay emphasis on the affix or word building teaching . Third ,the textbooks also should design more exercises on words building ,which is conducive to making the students master the word building better .

By following the three above approaches ,they can not only reduce the difficulties and obstacles in reading , but also enlarge the student s vocabulary . As a result ,it greatly improves the student s reading ability .

#### 4 ) Readabilities

Readability is known as the combination of structural and lexical difficulty of a passage .In fact ,lots of efforts have been made to make the readability of passage measurable ,and many formulas have worked ,such as the SMOGindex ,the Fry readability estimate ,the Flesh formula (qtd in Hou 2004 ) . The common purpose of these methods is to provide an easy and operable way to assess the difficulty level of the input passage . Frequently used readability formula is the Flesh ,first used in 1948 and still in use today . The formula produces a reading ease score . In this paper ,the Flesh formula is adopted for the calculation of the readability . The formula is like this : $RE4 = 206.835 (0.846 \times NSYLL) (1.015 \times W / S)$  (qtd in Alderson 2000 71 ) . On the grounds of the Flesh formula ,the readability of reading passages in 11 papers is calculated by computer and described according to readability yardstick as follows :

Table 4 .4 Readability of reading passages from 2003 -2007

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	2003
R	55 .5	70 .5	65 .2	62 .2	64 .2	55 .2	60 .8	64 .2	52 .7	64 .9	44 .7
D	Fairly difficult	Fairly easy	Standard	Standard	Standard	Fairly difficult	Standard	Standard	Fairly difficult	Standard	Difficult

R :Readability ;D :Description of difficulty

Table 4 .5 Readability classifications of reading passages from 2003 to 2007

	30 -50 difficult		50 -60 Fairly difficult		60 -70 standard		70 -80 Fairly easy		80 -90 easy	
Total Passages	#	%	#	%	#	%	#	%	#	%
55	15	27 .3	8	14 .5	19	34 .5	11	20 .0	2	3 .6

Through the comparison from the tables shown above ,we see that out of eleven papers within the

REscale (with forty as its low and seventy as its high) corresponds to the level of between "difficult" and "standard" according to the Flesh readability standard. And readability within the scale exactly responds to the reading ability of the students in the seventh or eighth grade and high schools in the U.S. Native speakers at this level cover 25% to 27% of the adult population. As far as reading passages are concerned, readabilities for different levels are involved (See Table 4.5) and beneficial to both the different learning levels of students stays away from discriminating against the excellent students and poor students. Therefore, a conclusion can be drawn based on the analysis that most papers get a good control of difficulty of reading passages from the perspective of readability.

However, readabilities below fifty take up 27.3% of the total, which corresponds to the level "difficult" according to the Flesh readability yardstick and responds to the reading ability of the students in high schools in the U.S. Moreover, the native speakers at this level just occupy about 24% of the adult population. It can be explained that they might be of much difficulty for the readers and probably exceed the desirable readability level prescribed by the syllabus. In contrast to the other papers, the readabilities of all reading passages in Paper 2003 are below fifty, which indicate obviously Paper 2003 is more difficult than the other papers. Of course, Yang and Weir (1998) states that it should be noted that the formula is designed for the English native learners not for the foreigners, but it can be regarded as a reference for the test designers and provide us a basis to build our comparison with a uniform criterion.

## 5) Topics

The teaching syllabus (2000) stipulates that senior middle school graduates can read and understand some materials on such topics as biography, story, popular science and modern technology, social life and culture, literature and history. Moreover, the New Curricular Standards (2001) describe 24 topics including personal information, history and geography, society and so on. Based on the requirements of teaching syllabus and the New Curricular Standards, topics will be classified in Figure 4.1.

S & C :Social life and Culture ;S :Story ;E :Education ;P S & M T :Popular Science and Modern Technology ;H & G :History and Geography ;T :Travel ;B :Biography

Figure 4.1 Topics of reading passages from 2003 to 2007

Alderson (2000) suggests that good tests of reading and good assessment procedures in general will ensure that readers have been assessed for their ability to understand texts in a range of different topics. As Weir (1993: 67) states, ideally we need to select the topic from a suitable genre, at an appropriate level of specificity, and the topic should not be culturally biased or favor one section of the test population. By analyzing the chart above and test papers, the writer finds out that the topics of reading materials embodies some characteristics as follows:

First, the selection of topics is extensive. Generally speaking, the reading comprehension parts of the eleven papers have a very broad selection of topics which almost cover all the topics required by the teaching syllabus. A great variety of selections of topics are conducive to reducing the bias of reading materials which is beneficial to some candidates. As a result, it can assess the candidates' reading abilities more scientifically, objectively and thoroughly.

Second ,different topics account for different proportions . Among of them ,sociallife &culture and popular science & modern technology accounts for a bigger proportion of 65 % . Whereas ,history & geography travel ,biography ,story and education occupy a smaller proportion .It can be explained that sociallife &culture and popular science &modern technology belong to general topic which contain some familiar ,necessary background information for most test takers ;and the rest are the general topic which is somewhat unfamiliar ,academic or technical to some individuals ( Hou 2004 ) . Bachman (1990 138 ) put forwards that “if the subject matter of input is familiar to some test takers and not others ,these individuals may have an unfair advantage ,resulting in better performance ” . Meanwhile ,Young and Weir (1998 ) also stress that “ideally ,the topic should be generally accessible ,i. e. not too obscure and not too familiar ,to all candidates ” . Therefore ,different topics in reading tests accounting for different proportions are reasonable .

Third ,most topics are quite novel and vivid . For example ,Passage A in 2006 ( 1 ) introduces greyhounds about its physical characteristics ,behaviors and traits and so on ;Passage A in 2005 ( 1 ) mentions a phenomena that pets owners are encouraged to take their animals to work and so on . Own to the novelty and interest of most topics ,it can greatly stimulate the readers interest in reading ,which is beneficial to make the test takers perform better in reading tests .

Fourth ,some topics are instructive and moral . For example ,Passage D in 2006 ( 2 ) narrates the significance of friendship and teach people how to deal with the problems about friendship ;Passage C in 2004 ( 1 ) talks about forgiveness and the author proposes that forgiveness can be surprisingly helpful to your physical and mental health ;Passage A in 2004 ( 1 ) involves protecting gorillas . These reading passages cannot only enrich the students visions ,but only develop their virtues and teach them how to get on well with the other people and how to do with some troubles in the life .

Finally ,most topics are authentic . Most topics of the reading passages are very close to the students daily life or people s modern social life . For example ,2004 ( 2 ) talks about the party custom . As for different age groups ,there are different customs of attending the party . Such topics make the test takers fell familiar ,which can actuate the readers prior experience or background knowledge and contribute to the comprehension process .

6 ) Genres

As for style ,it is used here in a similar sense to what Bachman (1990 138 ) called “genre ” . Young and Weir (1998 ) describe the genre from narration ,argumentation and exposition . However ,the teaching syllabus (2000 ) stipulates that besides narration ,argument and exposition ,the students should read and understand some simple practical writing such as ,letters ,invitations ,notices ,notes and some common signs . As a result ,on the basis of reasons above ,the writer of the paper tends to analyze the stylistic characteristics from narration ,argumentation and exposition and practical writing . A survey of the genres of the passages in these eleven papers is presented as follows :

Table 4 .6 Genre distribution of reading passages from 2003 to 2007

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	03	% total
N	3	2	2	2	2	2	0	2	2	3	0	36.4
E	2	1	1	1	0	1	2	2	1	1	1	23.6
A	0	2	1	2	2	2	2	1	1	1	3	30.9
P	0	0	1	0	1	0	1	0	1	0	1	9.0

N :Narration ;E :Exposition ;A :Argumentation ;P :Practical Writing

From the table above ,a thorough look suggests that genre from 2003 to 2007 can be described as follow :

Generally speaking ,the genre of reading materials is very extensive . Each genre can find its place in

reading tests ,which almost meet the requirements for genre in the syllabus .

Narration ,expository statements ,and argumentation account for different proportion ,36 .4 % , 23.6 % ,and 30 .9 % respectively . Alderson (2000 ) suggests that narrative texts are easier to process than expository texts ,perhaps because of the greater variety of relationships among text units ,possibly due to greater variety of content . As for exposition and argumentation ,they both pose a higher requirement on readers in that they either objectively elaborate on a certain theory ,method ,and fact or try to be persuasive enough to bring the readers to their viewpoints ( Hou 2004 ) . Therefore ,as for Senior Three graduates ,the proportion for narration and exposition is proper and reasonable . However ,the proportion of 30 .9 % for argumentation could be a little bigger .

It should be noted that practical writing with the smallest proportion of 9.0 % includes some advertisements ,but does not cover letters notice ,invitations ,notices ,notes and some common signs and so on which are required by the teaching syllabus . Therefore ,it can be explained that it can not meet the teaching syllabus well . So a possible suggestion can be made that more balanced distribution of every genre should be more concerned in the future tests ,which will promote English teaching and learning .

#### 4.2 Characteristics of Expected Response

Expected response is the “language use ” or “physical response ” test that attempts to elicit by the instruction ,the task ,and the input provided . The characteristics of the expected response are analyzed mainly from response types and reading skills .

##### 1 ) Response type

The types of response include the “selected response ” ,“limited production response ” and the “extended production response ” . The multiple choice questions ( MCQ ) format ,as a selected response type ,is used widely in reading testing ,including NMET ,CET 4 ,CET 6 and two major international language tests TOEFL and GRE . Due to its objectivity of answers ,MCQ can be scored by machine ,which ensures that marking is rapid ,entirely objective and reliable . In addition ,a series of statistical tools are available for item analysis ,test analysis ,score equating ,item banking of selected response items ,tests of this type can be pre tested fairly easily ( Weir 1990 ) . However ,there are some demerits of MCQ format measuring reading ability . As Weir (1990 44 ) argues ,“answering multiple choice items is an unrealistic task ,asin the real life one is rarely presented with several alternatives from which to make a choice to signal understanding ” . Moreover ,MCQ format allows guessing . Statistically a reader can get 25 % right ratio by guessing in each item .

By making a careful examination of reading tests ,the writer discerns that almost all selected response or multiple choice questions ( MCQ ) format has been utilized in the reading comprehension tests of NMET in recent years . Therefore ,the author makes a proposal that reading test formats are various as much as possible in the NMET reading tests ,such as short answer questions ,information transfer .

##### 2 ) Reading Skills

Alderson (2000 ) points out that the act of reading consists of the deployment of a range of separate skills ,abilities or strategies . Almost every armchair speculator seems to come up with a somewhat different list . However ,there are six reading skills described in the testing syllabus of NMET . Specifically ,they can be described as follows :

- 01 understanding the purpose and gist
- 02 understanding explicitly stated information
- 03 deducing unfamiliar words and phrases through context clues
- 04 making simple judgment and inference
- 05 understanding the basic structure of passages
- 06 understanding a writer's intention and attitude

According to the six reading skills listed above ,the distribution of skills in the reading tests of NMET are presented as follows :

Table 4 .7 Reading skills coverage of reading tests from 2003 to 2007

	07 (1)	07 (2)	06 (1)	06 (2)	05 (1)	05 (2)	04 (1)	04 (2)	04 (3)	04 (4)	03	% total
01	1	2	3	2	2	2	1	2	2	2	2	9 .5
02	9	8	8	8	10	7	11	10	10	13	13	48 .6
03	1	2	1	1	3	1	2	2	0	0	2	6 .8
04	8	8	6	8	4	7	4	5	5	1	2	26 .4
05	0	0	1	1	0	1	0	0	2	2	0	3 .2
06	0	0	1	0	1	2	2	1	1	2	1	5 .0

Great importance has been attached to the coverage of reading skills in the assessment of any reading test . This is no exception . Judging from the data collected , the following conclusions can be reached :

Various as their proportions are , the specific reading skills labeled as 01 ,02 ,03 ,04 ,05 and 06 can almost find their places in the reading comprehension parts of the recent five years .

The eleven test papers tapping micro skill 02 have a ratio of 48 .6 % , which covers the biggest proportion of the total items . It suggests that the understanding of the surface or literal input can exert great impact on the total score . It is very critical to a complete success of the reading test . It can be explained that micro skill 02 belongs to the lower order skills and is beheld as an elementary requirement for reading comprehension .

Micro skill 04 covers a bigger proportion of 26 .4 % , which suggests that it is impossible for test takers who just understand the local or surface information to get high marks . Test takers need such higher-order skills as judgment , inference and induction and so on to improve their reading comprehension abilities .

As shown from the above table , it is obvious that micro skill 05 covers a very small proportion in these papers , only 3 .2 % . Moreover , half of the test papers do not involve the skills . Therefore , in the coming reading tests , the test developer should try to take full into consideration on every reading skill ; or the micro skill 05 should not be required in the testing syllabus .

As a result , the analysis from above indicates that the reading comprehension component from the eleven papers have all exerted a good control over the proportion and distribution of the reading skills stipulated in the testing syllabus of NMET .

## 5 . Conclusions and Suggestions

### 5 .1 Conclusions

Through the detailed analysis of a considerable amount of data , the reading comprehension tests of NMET in recent five years embody the following characteristics :

First , the reading test papers from 2003 -2007 remain steady well for the test structure and format . Specifically , all the reading tests of NMET consist of five passages with 20 questions that are required to process in 35 minutes and primarily adopt the MCQ format to construct the reading questions .

Second , the eleven papers all exercise well control of the input capacity and reading speed , and they well answer the teaching syllabus and the testing syllabus .

Third , the topic of reading materials is full of wide extensiveness , novelty , education , morality and authenticity .

Fourth , most reading materials are selected with appropriate difficulty . Vocabulary and readability can be regarded as two most important factors to estimate difficulty of reading materials .

Fifth, reading skills gradually emphasizes the examination of the textual understanding and integrated abilities.

Therefore, the reading tests of NMET from 2003 to 2007 have well met the requirements described by the two syllabi and have reached satisfactory content validity.

## 5.2 Suggestions

Although reading comprehension part of NMET has a high content validity, nothing is perfect. Based on the present research, some suggestions are put forward with a purpose for further improvement and development of the reading tests in the future.

First, test format should be enriched.

The author discerns that the reading comprehension tests in the recent years only adopt multiple-choice format. Weir (1993) indicates that multiple choice tests exhibit almost complete marker reliability as well as being rapid and often cost less to mark than other forms of written test. However, there are some demerits of MCQ format measuring reading ability, such as allow guessing; place an unnatural burden on the test takers. It is based on this very concern that the author suggests that test formats should be enriched in the reading test, such as short answer questions, information transfer.

Second, reading speed should be slightly raised.

Through making the comparison of two reading speeds, the author finds out that the reading speeds just reach the first grade, not reach the second grade in the teaching syllabus. However, NMET, as a power and speed test, it mainly aims at selecting the qualified candidates at the university level. Therefore, the author suggests that reading speed of NMET needs to be slightly increased.

Third, the difficulty of reading materials should be under better control.

Readability is regarded as a significant factor measuring the difficulty of the reading materials. According to the Flesher readability standard, readability of reading materials between fifty and eighty is appropriate for Senior Three students. However, the readabilities of some reading materials are below fifty, taking up 27.3% of the total, which may be much difficult for the Senior Three graduates. Therefore, the author suggests that, the readability of reading materials should be under better control.

Fourth, the layout of reading passage should be arranged appropriately.

Generally, reading passage should be arranged sequentially according to the principle "from easy to difficult", or "from short to long". However, the author discerns that there are some weaknesses to the layout of reading passages. The layout of the future reading passages should be arranged appropriately to make sure test takers have appropriate time to mentally prepare for the task at hand.

Fifth, the text should be presented with its title.

Through careful observation, the author of the paper detects that almost all of the reading passages are presented without their titles. It is well known that the title of the reading passage helps the reader understand the content of the passage. Most reading materials such as newspapers and magazines all contain the titles which retain the original form for information presented in the text. Therefore, the author suggests that each passage should contain a title.

Sixth, the students should strengthen and widen reading.

In NMET, reading comprehension accounts for approximately 27% of the total score, which has been regarded as a decisive factor affecting the testees success in NMET. Therefore, in the regular teaching and learning, students should pay close attention to reading. In addition, the reading passages in the eleven papers have a very broad selection of topics, including social life and culture, popular science and modern technology, education, story and travel and so on. The job of these papers is to transmit information to us, the reader. Students should read on a regular basis to broaden their academic horizons and enrich their personal lives. In turn, this will improve their reading comprehension and fluency.

Seventh ,word building should be emphasized in the regular teaching and learning .

In the analysis of new words ,the writer finds out a fact that about 80 % of the new words are compound and derivative words . Therefore ,in the regular teaching ,the teachers should lay more emphases on the word building . Textbooks should include more exercises that focus on word building , which is conducive to expanding students vocabulary .It can reduce difficulties and obstacles that may stand in the way of a struggling reader .

### 5.3 Limitations of the Study

Although great efforts have been made in the study of reading comprehension parts of NMET ,there still exist some weaknesses .It is insufficient to depict the full picture of the reading comprehension parts of NMET due to the lack of the testees performance data and the investigation for the testees and testers . Only the reading comprehension parts of NMET national paper are involved in the present research ;resulting in an incomplete depiction and understanding of the NMET .

### References

- Alderson ,J . C . 2000 . Assessing Reading . Cambridge :Cambridge University Press .
- Alderson ,J . C . , C . Clapham & D . Wall . 2000 . Language Testing 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 . 1991 . What Does Language Testing Have to Offer ? TESOL Quarterly 25 /4 .
- Bachman ,L . F . & A . S . Palmer . 1999 . Language Testing in Practice . Shanghai :Shanghai Foreign Language Education Press .
- Henning ,G . 2001 A Guide to Language Testing : Development , Evaluation and Research . Beijing :Foreign Language Teaching and Research Press .
- Hou ,Yanping . 2004 . Assessing reading by a model of task characteristics in content validation —Evaluating the reading part of TEM4 ( 2000-2002 ) . Unpublished M . A thesis , Shanghai International Studies University .
- Hughes ,A . 1989 . Testing for Language Teachers . Cambridge :Cambridge University Press .
- Nuttall ,C . 1982 . Teaching Reading Skills in a Foreign Language ( 1st edition ) . London : Heinemann Educational Books Ltd .
- Qi ,Luxia . 2005 . Stakeholders conflicting aims undermine the washback function of a high stakes test . Language Testing .
- Weir ,C . J . 1990 . Communicative Language Testing . London :Prentice Hall International ( UK ) Ltd .
- Weir C . J . 1993 . Understanding and Developing Language Tests . Hemel Hempstead :Prentice Hall International ( UK ) Ltd . .
- Zou ,Shen . 1997 . The test for English Majors ( TEM ) Validation Study . Shanghai :Shanghai Foreign Language Education Press .
- 刘润清、韩宝成 2004 《语言测试和它的方法》。北京 :外语教学与研究出版社。
- 《普通高等学校招生全国统一考试大纲的说明(英语)》,2003-2007 ,北京 :高等教育出版社。
- 《全日制普通高级中学英语教学大纲(实验修订版)》,2000 ,北京 :人民教育出版社。
- 许之所、张丽芳 ,2004 ,高考英语试卷预测效度实证研究,《武汉理工大学学报(社会科学版)》第2 期。
- 杨惠中 ,C . J . Weir 1998 《大学英语四、六级考试效度研究》。上海 :上海外语教育出版社。
- 《英语课程标准(实验稿)》,2001 ,北京 :北京师范大学出版社。