



## National Matriculation English Test in China: Its Past, Present and Future

**Deliang Wang\***

Beijing Normal University, Beijing, China

**Xijia Zhang**

University of Alberta, Edmonton, Canada

### Abstract

The National Matriculation English Test in China is the most high-stake exam with the largest number of test-takers in the world. This paper explores the past, present and future of this test. Because of the specific social and political circumstances, this test started to take forms in 1977. Based on the analysis of the format and question types of all the test papers, we find that this test has gone three stages. Each stage has its own characteristics and was influenced by different educational theories. In 2014, the State Council launched a radical reform, which was considered as a new starting point for this test and will definitely exert wide-spread impacts on related parties. By reviewing the history, analyzing the present and forecasting the future, we can have a better understanding of the matriculation test.

**Key words:** matriculation test; English test reform; high-stake exam; historical changes

### INTRODUCTION

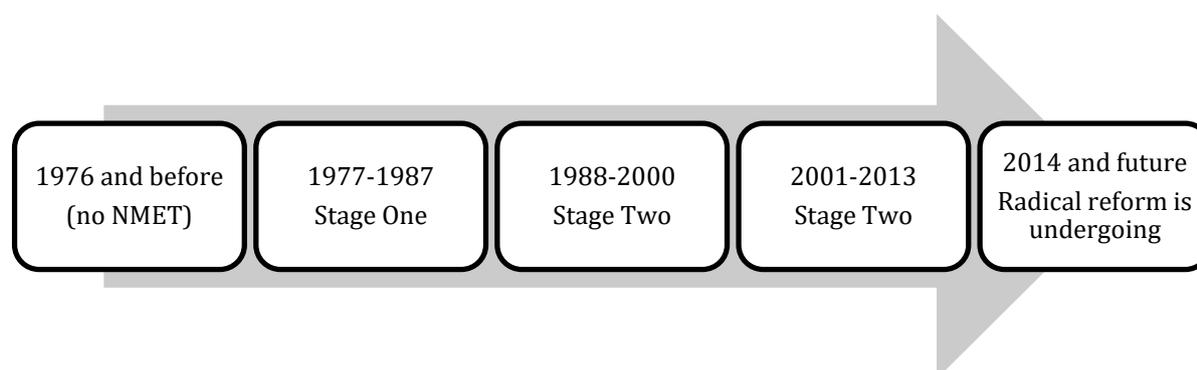
Ever since the end of the Cultural Revolution, reforms have been taking place in different fields in mainland China, including the English assessment and testing system (Shao & Huang, 2006). Before the year of 1977, due to the Cold War with the western countries, English testing and assessment in mainland China had always been ignored. During this period of time, the design of the National Matriculation English Test (NMET for short) was relatively casual. Students were even allowed to use Chinese to answer some of the questions (Liu, 2005). Nevertheless, the formulation of the opening-up and economic reform policy asserted a great influence on the field of education (Wu, 2013), especially on English testing and assessment system (Liu, 2005). Since the communication with other foreign countries, especially English speaking countries, would become more frequent after the policy was put forward, scholars began to realize that it was important to develop a comprehensive and effective assessment system of students' English ability among the nation (Cheng & Curtis, 2010). Reforms that take place in NMET usually reflect changes in other significant social fields, like education, philosophy, and even politics. For example, in the 1960s and 1970s when the diplomatic relations between China and the

Soviet Union turned cold and the normalization of relationships between China and the U.S., the focus of language education in mainland China also turned from Russian to English. Accordingly, reforms in NMET can be seen as an essential social asset which should be recorded and remembered. They are part of the Chinese history, which are worthy to be remembered by future generations.

According to the database of CNKI (China National Knowledge Infrastructure)<sup>1</sup> from 1981 to 2015, the previous researches about NMET are neither complete, nor systemic and it is necessary to conduct further researches about all those reforms and changes that have taken place in NMET during the past thirty years. This is what makes this research significant. This research will focus on the history as well as the present situation of the reform and try to make some general predictions for its future path in order to shed some light on the development of English assessment system in China. All the test papers used in the history of NMET will be adopted as the data for the present study.

### HISTORICAL CHANGES OF NMET

As China is becoming more and more involved in international affairs, the society also puts forward different requirements for students' English abilities. The assessment and testing of students' English proficiency have been changing responding to these different requirements. In light of analyzing all the test papers collected as data, it is found that the development of NMET has gone through three stages in its history, that is, Stage one (1977-1987), Stage two (1988-2000) and Stage three (2001-2013). Because before 1977, China was undergoing the Cultural Revolution, the whole society fell into a chaos and there was no NMET in its true sense. On the other side, 2014 saw such a radical change that we have to put it on a completely new page. Therefore, the history of NMET falls mainly between 1977 and 2013, summarized as the following figure:



**Figure 1.** Historical Development of NMET

This division is made mainly based on the question types, test formats and constructs of testing design. This study will probe further into them one by one.

#### Stage One

The first stage in the history of NMET, from 1977 to 1987, was merely a process of exploration. During these ten years, NMET in mainland China actually changed quite

<sup>1</sup> www.cnki.net

frequently. It can be seen from the test papers of these ten years that a different type of question was added every year. What is interesting is that though question types varied annually, the testing objectives of these questions seldom changed.

As far as the question types and formats are concerned, a careful analysis revealed that they can be divided into four groups according to their divergent testing objectives. The first group mainly focused on testing students' pronunciation, such as *choosing the letters with the same or different pronunciation, classifying given words according to the vowels they include in their pronunciations*. Listed below are some of the examples of this particular group of questions:

1. Underline the word whose vowel pronounces different from others.

e.g. *be me wet need see*

(1). beat meat break heat freeze

(2). have gave wave save slave

(The 2<sup>nd</sup> question in 1982's test)

2. Classify the following words according to the different pronunciation of their vowels

e.g. [au] house now

*die field heard peace fly keep high work third know term key stone league boat turn*

(1). [ai] \_\_\_\_\_

(2). [i:] \_\_\_\_\_

(3). [ə:] \_\_\_\_\_

(4). [əu] \_\_\_\_\_

(The 1<sup>st</sup> question in 1983's test)

The second group was primarily concerned with vocabulary. Questions in this group often asked students to select the correct English explanations of certain words, or to use synonyms to replace the original words without changing the meaning of the sentences.

3. Match the words with their corresponding English explanations

(1). alive ( ) at once

(2). consider ( ) order somebody not to do

(3). discover ( ) take away

(4). enter ( ) not dead

(5). feed ( ) take place

(The 3<sup>rd</sup> question in 1984's test)

4. Choose the synonyms of the given words from the chart to fill out the blanks without changing the meaning of the entire passage.

rise	found	returned	arrive at	have	became
married	being wounded	leave	almost	(was paid)	

“When I was your age,” said the old man, “I was paid only \$30 a week. Life was much harder then. I had to \_\_\_\_\_ at six and \_\_\_\_\_ the factory by seven. We worked ten hours a day for six get up days a week and didn’t \_\_\_\_\_ any holidays.

(3) get

(Section B of the 3<sup>rd</sup> question in 1985’s test)

The third group was about grammar, including question types such as *choosing the correct word to finish the sentence, using the correct form of the given words to fill in the blanks, using the correct tenses of the verbs to finish the sentences* and so on.

5. Multiple Choice: Choose the correct answer to fill in the bracket

(1). John is ( ) university student.

(A) some (B)any (C)a (D)an

(2). That table is made of ( ).

(A) a wood (B) some wood (C)the wood (D) wood

(3). Tom ( ) a letter to his parents last night.

(A) writes (B)wrote (C) write (D) has written

.....

(The 1<sup>st</sup> question in 1979’s test)

6. Use the correct tense of the given words to fill in the blanks

Thousands of years ago, the Egyptians believed strongly in life after death. They thought a person \_\_1\_\_ (need) his body in an afterlife. Therefore they treated the dead body with spices (香料) and oil and then wrapped it in cloth. The\_\_ 2\_\_ (wrap) body was then put in a tomb. A body \_\_3\_\_ (treat) in such a way \_\_4\_\_ (call) a mummy.

.....

(The 4<sup>th</sup> question in 1986’s test)

The fourth group contained questions about how to utilize some basic English skills to solve problems. Question types ranged from *cloze tests*, *reading comprehension*, and *dialogue completion* were designed to test students' comprehensive English proficiency. *cloze tests*, and *reading comprehension* are common question types, so we only give an example of *dialogue completion*.

### (7) Dialogue Completion

假设你是售货员。有个名叫李培的顾客来书店，花了一元钱买了两本《怎样保护视力》的书。下面是买书时的一段对话。请把你所说的五句话补全。打问号的，必须用疑问句；打句号的，必须用陈述句。

#### At the Bookstore

Assistant: 1 ?

Li Pei: Have you got anything that tells people how to take care of their eyes?

Assistant: Yes. 2 ?

Li Pei: Good. How much is it?

Assistant: 3

Li Pei: That's cheap enough.

Assistant: 4

Li Pei: I'll take two copies. How much is that?

Assistant: 5

Li Pei: Here's five yuan.

Assistant: Thank you. Here's your change—four yuan.

(The 7<sup>th</sup> question in 1986's test)

Judging from the question types in these ten years, it is clear that NMET from 1977 to 1987 had some basic characteristics:

#### ***Ignorance of students' actual English proficiency***

To begin with, the tests paid more attention to students' basic English abilities like the vocabulary and the grammar, while the importance of comprehension and general English performance were largely understated. In all ten test papers, over 50% of the question types were about the memorization of words and grammatical rules, which served as a sound evidence for the emphasis on vocabulary and grammar.

On the other hand, in terms of testing students' English language performance, three other significant abilities, that is, listening, speaking and writing tasks, were never seen. Though there were questions about how to distinguish different sounds, there was no question types developed to test students' listening ability. Since seeing the transcription

of these words was totally different from hearing them in the recording. In this way, it may be concluded that the listening ability of students had been ignored to some extent. In terms of writing ability, matriculation tests in these ten years included questions like paraphrases and dialogue completion, which were used as a simple way to test students' English writing ability. Tasks like comprehensive writing were not designed, which served to prove that English education at that time might not pay much attention to students' aptitude to use English to convey a full message. As for speaking, in these ten years' tests, there was no oral English examination. Students were merely asked to remember the phonetic transcriptions of the English, just as how they were requested to memorize the spelling of the vocabulary and the grammatical rules, which was far from enough for the training of their speaking English. Without a separate oral English test, it would be extremely difficult to assess students' oral ability. Thus, the overemphasis on basic knowledge, together with the ignorance of students' real proficiency to use English in real life, was the most conspicuous feature of this period of testing. In addition, in terms of grammar, tests in these years concentrated more on lexical rather than syntactic grammatical rules. Nearly 80% of the grammatical question types in these ten years' tests were about things like verb tenses, modal verbs, and so on. Syntactic rules were examined in a quite simple way. Compound sentences like attributive clauses, nominal clauses and adverbial clauses were excluded from the question types.

Thirdly, there were usually bonus questions designed for students whose level are high above the average, providing a possible way to mark the differences among students' English aptitude from these extra tasks.

### ***Reading for the sake of reading***

As for reading ability, it had received more and more attentions in these ten years. Before 1985, there was only one reading assignment. However, it increased to three by 1987, which indicated that more emphasizes were put on students' proficiency to read English articles. Meanwhile, for these reading questions, most of their answers could be found directly from the reading paragraphs. Questions about the main ideas of these articles, as well as the attitude of the authors, seldom appeared in the test papers, with only three questions in ten years' time. This meant that students at that time were only required to identify the specific details of English articles but they were not asked to train their ability to summarize the main ideas and the general attitudes of these articles, nor to make inferences.

### ***Behaviorism as the rationale of the reform in this stage***

It is evident that the scholars and English teachers of mainland China did not design all the above question types randomly. As a matter of fact, the application of these question types and formats is related to one of the most popular psychological theories—behaviorism—during that period of time. It emphasizes that it is the observable behavior of people and animals, not unobservable events that take place in their minds, which matters most (Skinner, 1984). Behaviors as such can be described scientifically without recourse either to internal physiological events or to hypothetical constructs such as thoughts and beliefs (Baum, 1994).

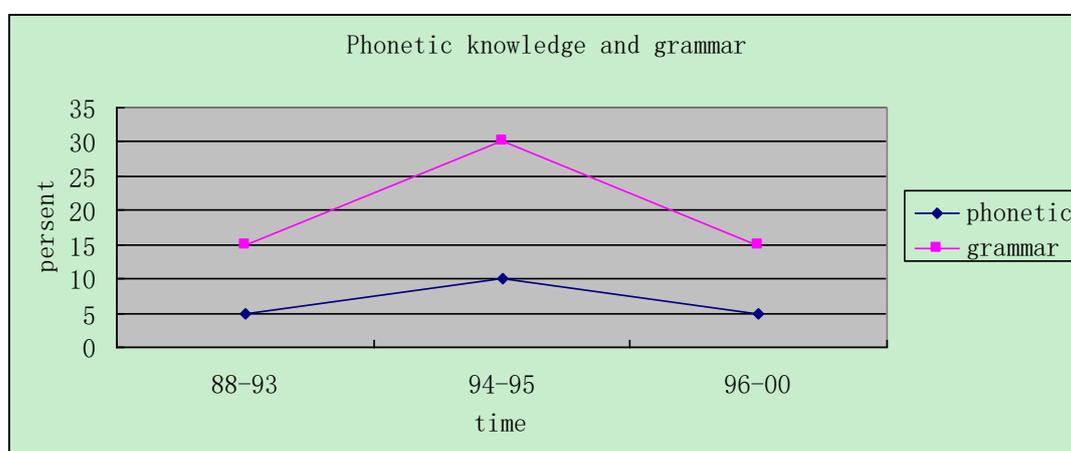
When it comes to education, suppose students get good grades in their tests and their parents and teachers praise them, students would feel very happy. The success of the tests may be a neutral stimulus, but the students gradually connect it with the praise of parents and teachers, which is a conditional stimulus. As time goes by, the success in the tests may cause the students to be happy. Hence, this theory is adopted by many educators and teachers. They try to emphasize the importance of getting good grades by providing all kinds of positive stimulus after the students' excellent performance in the tests, which lead to the test-oriented atmosphere in English education in China to a large extent. According to Skinner (1984), if teachers reward students after some of their favorable reactions, like studying hard or responding actively in the classes, these behaviors, or reactions, would be reinforced and appear more often, while those behaviors that are punished will decrease. That is the methods constantly employed in English teaching in China. As a matter of fact, Skinner's theory has always been recognized as the most influential theoretical framework in the education system in China in 1970s and 1980s.

### Stage Two

The second stage starts from the year 1988 and lasts for a dozen years. During this time, as the social development gradually stabilized, NMET also took up a more settled form. Instead of constantly adding new question types every year, the formats varied little during these years. However, though the focus of the test gradually turned from memorizing existing knowledge to evaluating students' English aptitude in a more thorough way, the major concentration was still about memorization.

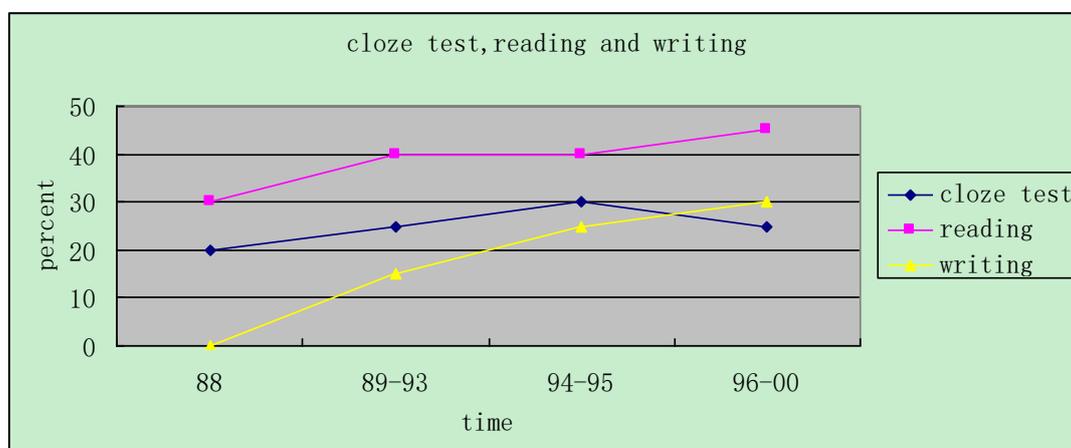
The test format during this stage was more coherent compared with the previous ten years. There are generally six divergent types of questions: *phonetic knowledge*, *grammatical questions*, *cloze test*, *reading comprehension*, and *writing test*. In the meantime, in 1988, there was no writing test while from the year 1996 to 2000, there was a test of vocabulary.

First of all, in terms of phonetic knowledge and grammatical questions, their percentage in the whole test varied quite often during this stage, which can be illustrated by the following figure:



**Figure 2.** Historical Changes of Phonetic Knowledge and Grammar

Secondly, during this time, NMET put more emphasis on examining students' practical English abilities. The appearance of the new question types, *cloze test* and *writing test*, is the best embodiment. Since 1988, the percentage of cloze test saw a gradual increase. So did it with the reading comprehension and the writing test.



**Figure 3.** Historical Changes of Cloze Test, Reading and Writing

In addition, in 1994, a new kind of question, *error correction*, was designed to test students' comprehensive English aptitude. This is a comprehensive question type since when designing the errors, teachers and scholars can put spelling errors, grammatical mistakes, and coherence problems in this single question type. In this way, students have to employ their overall English proficiency to get a satisfactory score in this question.

Judging from the above, it is obvious that the English language education in China during these twelve years began to pay more attention to the development of students' overall English ability. Between 1988 and 2000, unlike the previous ten years, NMET was more mature in both testing format and content. The designers of the test started to pay more attention to the assessment of students' general English abilities rather than their memorization of basic knowledge. That is why there were an increased proportion of comprehensive question types, such as *cloze test*, *reading comprehension* and *writing task*. Furthermore, there was also a new question type, *error correction*, which served to test students' language use ability. In terms of students' reading ability, they got more opportunities to summarize the main ideas of the reading paragraphs. They were also offered more questions which required them to use their inference ability to get the right answers.

Another vital characteristic of this stage is the uncertainty of the different weights of each part of the test. From Figure 3, it is shown that the percentage of basic knowledge tends to fluctuate during this period. The weight of phonetic knowledge went from five points in 1988 to 1993, to ten points in 1994 to 1995, and then decreased to 5 points again in 1996 to 2000. So is the weight of grammatical questions. As for vocabulary, the situation got more extreme. In 1988, there was a 10-point vocabulary question. Then, from 1989 to 1995, this question type was eliminated from the whole test before it appeared again in 1996. These mean that scholars and teachers at this time were still not quite sure about the importance of different parts of English language education. They constantly changed

the weights of different question types since they did not have a clear idea about what the right proportion was.

Though there had been great progress in this stage of the reform of the matriculation test, some problems remained to be solved. For instance, the testing of students' listening and speaking ability was not included in the reform of this period. In this way, students' overall English proficiency could not be trained and developed to the full extent. Also, the direction of English education in this stage was still test-oriented rather than focusing on the development of students' real language competence.

During this period of reform, it was found that the behaviorism theory actually had a lot of defects and thus could not be used to explain every aspect of education process. For example, from the behaviorists' point of view, "teachers" should take complete control in educational activities. In this way, learners do not have any opportunity for evaluation or reflection within the learning process. They are simply told what is right or wrong. The conceptualization of learning using this approach could be considered "superficial" as the focus is on changes in external behavior instead of the internal processes of learning leading to behavior change (Friesen, 2005). Also, another most notable criticism toward behaviorism was Chomsky's argument that language could not be acquired purely through conditioning, and must be at least partly explained by the existence of internal mental states (Chomsky & Skinner, 1959).

The reflection of the behaviorism gave rise to another important educational psychological theory—cognitive theory. In the second half of the 20th century, behaviorism was largely eclipsed as a result of the cognitive revolution (Madden, 2013). Instead of seeing the process of learning simply as a stimulation-reaction phenomenon, the cognitive theory pays more attention to the underlying psychological activities during the time of study. Scholars of this theory believe that the way people think impacts their behavior and therefore cannot be a behavior in and of itself, as what the behaviorists have claimed. They later argued that thinking is so essential to psychology that the study of thinking should become its own field (Lilienfeld, et. al. 2010). In this system, people's learning activity is divided into different procedures which take place one after another in a logical and coherent order.

Under the influence of this theory, teachers and scholars in mainland China started to emphasize students' cognitive process as well. It can be seen that all the question types from 1988 to 2000 were designed based on the logical order of students' cognitive process, that is, from memorization and recognition of basic knowledge, to the understanding of more complex situation, and finally to the comprehensive application of what has been learnt. Concretely speaking, in the real test, the question types started with simple ones like phonetic knowledge and grammatical rules which students simply had to recall what they had been taught in the classroom. After that, there was the cloze test and reading comprehension which aimed at testing students' ability to use what they had learnt including vocabulary and grammar to understand the meaning of these texts. Finally, at the end of the test, there was the question of applying what they had learnt to solve real world problems and to create something new, which was illustrated by the error correction and the writing task.

### Stage Three

The third stage of the reform went from 2001 to 2013. This is the most developed stage ever since English became a subject of the College Entrance Examination. Seen some trials and hardships, NMET has finally come to the stable and developed period. During these thirteen years, the test followed a more consistent format and the weights of different question types remained the same as well. Test paper used from 2001 to 2013 was composed of six question types: *listening comprehension*, *use of language*, *cloze test*, *reading comprehension*, *error correction* and *writing task*. Their formats and weights are shown in Table 1.

**Table 1.** Test Format in 2001-2013

	Format		weight
Part I	Listening Comprehension	Section 1 short dialogue	7.5
		Section 2 long dialogue or monologue	22.5
Part II	Use of language	Section 1 multiple choice	15
		Section 2 cloze test	30
Part III	Reading Comprehension	Choose the best answer	40
Part IV	Writing	Section 1 error correction	10
		Section 2 letter writing	25
Total			150

This is the very first time that *listening comprehension* was included in NMET. It contained two different parts: five short dialogues and five long dialogues or monologues, which weighted 20 percent in total.

*Multiple choice* mainly examined students' mastery of grammatical rules, and it also tested students whether it was decent to say certain things under particular circumstances, that is, the practical use of language. This part was meant to test students' pragmatic competence. All the questions were multiple choices, taking up 10% out of the total test.

As for *cloze test* and *reading comprehension*, their weights have increased again to 30 points and 40 points, while the points of error correction and writing task both decreased 5 points compared with the years from 1996 to 2000.

As is mentioned above, this stage witnessed one of the major breakthroughs in the reform of NMET in mainland China. Two important question types were employed to assess students' English proficiency, that is, *listening comprehension* and *use of language*. This means that the direction of the test was not simply test-oriented any more. The teachers and scholars in China started to attach more importance to students' ability to use English language in real life communication. Thus, they were required to be able to understand the information provided by others as well as to identify the possible implications in others' words, both of which had manifestations in the listening comprehension test. Also, the purpose of the appropriateness test was to evaluate their ability to understand the communication manners of English speaking people so that they would not say something offensive in real world communications.

The variation of question types in the reading test was another milestone in this stage. Before 2001, most reading question types were multiple choices, but starting from 2011, the last reading passage was always designed in a creative way. Students were either asked to complete the missing paragraphs or to put the paragraphs in a correct order. These kinds of creative reading tasks were more challenging yet more meaningful when it came to the assessment of students' reading comprehension competence.

The third major feature of this stage is that the test has finally started to stabilize. The question types and formats remained unchanged and so did the corresponding weights. This means that the scholars and the teachers have reached a mutual understanding of the respective status of different parts of English language education. Thus, they were able to use the weights and proportions in the test to show this understanding, which is a great progress in the history of English language education in China.

Finally, the only regret remained is the omission of the speaking test. Though in some parts of China, especially the eastern coast where the economic and social development are more advanced and the communication with foreign countries is more frequent, the speaking test has already been employed as part of the matriculation test, its score has not been taken into account yet and it has not received deserved attentions as well.

During this period of time, the development of cognitive theory became more mature, and went further in defining learners' essential position in learning activities. Scientists started to realize that learners' psychological activities mattered not only because they were triggered by study materials, but also because they were evidences of learners' active participation in the learning process. Thus the constructivism theory began to gain its popularity. Generally speaking, its major point of view is that humans generate knowledge and meaning from an interaction between their experiences and their ideas (Piaget, 1967). It exerted great impact on the reform of matriculation test in China in this period.

For one thing, the individual constructivism actually turned Chinese English education from teacher-oriented to more learner-focused. Instead of having teachers telling students everything they need to know, students were asked to explore by themselves more often. For example, they may be asked to summarize the structure of relative clauses by studying several typical ones. They may also be required to find out the logical connections among the reading texts by themselves. This kind of educational trend explains the appearance of the new question type in reading test to a large extent.

In the meantime, the appearance of *use of language* as well as *listening comprehension* showed that scholars in China began to realize the importance of developing students' English proficiency in a real life situation, since according to the constructivism, in real life, the surrounding environment will greatly affect the complexity of language communication. Hence, it is very important to have some simulations in English teaching and assessment so that students can get more chance to train their abilities to react properly to these situations. That also explains the origin of the communicative teaching methodology, as this teaching method always put a lot of emphasis on creating a real-life

setting for students to communicate in English, which accords to the perspectives of constructivism.

## THE PRESENT AND FUTURE OF NMET

### 2014 Reform- A Milestone

On September 3, 2014, the State Council issued the "Opinions on Deepening the Reform of the Examination System of Admissions", which became the headline of all mainstream media that day, because matriculation exams are the great concern of the public and they may affect a destiny of a student and his family. According to *the Opinion*, NMET will be carried out from once to twice each year. This will provide more opportunity for students. They can choose the higher score to apply for their ideal universities. "Opinions" shows that the new exam system will be conducted at some selected provinces in 2015. Up to 2017, the new examination system will be promoted comprehensively to all provinces in mainland China. In 2020, China will establish the modern education system of examination with Chinese characteristics, forming new examination and enrollment patterns with classification test, comprehensive evaluation, and multiple admissions.

The reformed NMET not only changes its frequency, but also its format and construct. And these changes are radical. They will be elaborated with details in the next section. From this sense, 2014 Reform will become a milestone in the history of NMET.

### New Test Format and Its Construct

Over the years between 2000 to 2014, the test format remained relatively stable. However, with the launching of 2014 reform, the test format will undergo a radical change. The new test format is shown in Table 2.

**Table 2.** Test Format in 2014 Reform

Format		weight	
Part I	Listening	Section 1 short dialogue	7.5
		Section 2 long dialogue or monologue	22.5
Part II	Reading	Section 1 reading comprehension	25
		Section 2 Complete a passage with sentences	10
Part III	Use of language	Section 1 cloze test	30
		Section 2 Filling in the blanks with words	15
Part IV	Writing	Section 1 practical writing	15
		Section 2 Continuation writing based on reading/summary writing	25
Total		150	

Compared with the previous test format, the new one has several obvious features. First, the biggest change is the writing part. Previously, the writing part includes mainly two parts: error correction in a passage and a practical writing task, usually letter writing. The difficulty level greatly increases with the new format. Error correction is completely abandoned from this year on. Error correction first appeared in 1991 and remained as one of the standing test types until now although it undergone some slight changes, such as its weight changed from 15 points to 10 points in 1996. Error correction has been controversial and whether it should be taken away has become one of the hot topics over

the past years. Some think that the test purpose of error correction is not clear and its test structure is not definite. It is difficult to realize the communication function as it claimed. Therefore, it is impossible for it to realize its assessment value as a language disciplinary test. However, others hold that error correction is a comprehensive test task and it can assess test-takers' evaluative ability of discourses and ability of accurate use of language (Yu & Huang, 2009). According to some empirical study, the wash back effect of error correction is very low. Qi (2007) does a questionnaire and interview investigation, finding that there is a gap between actual aftereffect and expected aftereffect. In many high schools, error correction-targeted training is widespread. Although training on error correction may help students to learn more grammar knowledge, it has little help in reading and writing ability development. The more training on error correction students receive, the more chances for them to face the input of wrong uses, which will affect students' improvement of their English level.

In the new test format, two alternative writing test tasks appear, that is, *continuation writing based on reading* and *summary writing*. These two tasks will pose a lot of challenges to students. Most of them have never done exercises like these. Examination Center of Ministry of Education has taken a very bold step in adopting these two tasks, because they are still under investigation and approval.

The test type of *continuation writing based on reading* was proposed by Chuming Wang. He and his colleagues have done a series of studies to testify its existence and adoption (Wang 2012, Wang and Qi 2013, Wang M. and Wang C. 2014, Wang C. and Wang M. 2014). The continuation task combines language input and output, intending to improve students' language use ability. The task is easy to design, as long as taking away the ending part of a reading passage and asking students to write the ending of their own version based on the reading comprehension of the left parts. Wang (2012) has testified theoretically and experimentally that the continuation task has some advantages: it can help improve students' creativity and imagination; setting up the context for language learning; imitating the standard language but creating new content; improve the ability of coherent expression by continuing the existing text; improve the efficiency of language learning by combining language comprehension and output; there are rarely negative influence upon language teaching and learning. Wang and Qi (2013) carry out an empirical study. The result shows that the learners' performance of the continuation task correlated significantly with their performance of the reading and writing tasks used in a standardized English proficiency test, suggesting that the continuation task is a valid measure of the learners' reading and writing abilities. The 2014 Reform adopts this test type, which is an official testification of this new test form, because it will be used in the most large-scale and high-stake exams in China.

Summary writing came to researchers' vision in the 1980s. Summary writing is a process that students reduce reading material in an original work to its main points and key supporting details. Students must "get inside" the material and fully realize what is being said before they can reduce its meaning to a few words. The ability to summarize information requires readers to sift through large units of text, differentiate important from unimportant ideas, and then synthesize those ideas and create a new coherent text

that stands for, by substantive criteria, the original (Dole, et. al. 1991). Recent years have witnessed that summary writing became more and more popular among large scale and high stake examinations abroad and home, such as TOFEL and The Matriculation Exam in Guangdong Province. That is because most of the studies report positive results about the effectiveness and reliability of this test task (Kirkland & Saunders. 1991, Robitaille and Connelly 2002, Weigle, S. C. 2004, etc.). This is also why 2014 Reform has taken summary writing as an alternative task.

The second biggest change is that the test type of multiple choices is taken away. Multiple choices used to be an indispensable part of Use of Language. It, as a discrete-point test, mainly tests students' knowledge of grammar and vocabulary. It is a product of structural linguistics. Although this test type is always disputed, it made a great contribution to Chinese Matriculation English Test. It appeared as one of the earliest and the most common test tasks since the beginning of modern matriculation English test in 1978. Multiple choice remained as an indispensable part over from 1978 to 2013, although it itself has undergone various minor changes, for example, it changes from knowledge-oriented to ability-oriented, from assessing grammar and vocabulary knowledge isolatedly to testing grammar, vocabulary and culture within certain contexts.

However, the test task of multiple choice has never escaped from criticisms from academia. Xiao and Luo (2002) hold that multiple choice lacks of construct validity, unable to assess test-takers' English level. Xu and Zou (2013) find that the multiple choice cannot assess test-takers' actual language ability, because grammar tests are inclined to assess grammar knowledge, hindering the development of language skills. Qi (2004) tends to be more radical. She proposed firmly that multiple choice should be completely deleted. She thinks that this question type completely ignores the communicative functions of language and gives the teachers in middle schools chances to separate grammar and vocabulary teaching from textbooks and drill their students mechanically. This test task is considered as the root of the sea of exams tactics and examination oriented education.

The 2014 Reform changed *multiple choice* into filling in the blanks with words. According to Xu and Zou (2013), the test type of filling in the blanks with words is more comprehensive and more distinguishable than multiple choice, because it not only takes into consideration the grammar knowledge, including form, structure and meaning, but also the proper application of these knowledge in a coherent text and real context. From this point, we can say that this reform takes a shift from isolated assessment to integrated assessment.

The third change is the way of organizing the exam. From 2015, the provinces chosen as the reform pilot will provide Matriculation test twice a year. Students can choose the higher score as the final score to apply to the university. The change is historical because none of the matriculation exams including all subjects has ever been given more than once a year in the history. This change is also very significant because it is more human-centered than people. In the past, if a test-taker makes some accidental and careless mistakes in the matriculation test, his or her fate will be influenced. We are humans, not machines. No one can guarantee they will never make mistakes especially under such a

psychological and physiological pressure. However, now if they are given a second chance, their basic human rights are protected. They have an opportunity to make up for what they think is not enough. Therefore, this change has received many applauds from various aspects since this news was announced.

### **Possible Impact on Future English Teaching in High School**

In China, there is a stubborn tradition in school education, that is, teaching is heavily exam-oriented. Teachers usually teach according to what are tested in the exams. Especially for exams such as the matriculation test, they are the baton of high school teaching. Usually, the matriculation test syllabus is carefully and heavily studied by all the teachers, and some experts or administration officials even join in explaining and interpreting this syllabus. Then teachers come back to their classrooms and teach accordingly. From this sense, the exam syllabus of 2014 Reform is strategically important and it will exert strong impact on the English teaching in high schools.

According to the 2014 Reform tenets, comprehensive English ability is given special attention. Therefore, middle schools will care more about students' comprehensive English ability, including speaking, listening, writing and reading. In the past, speaking and listening are always neglected because they are not the subjects of the matriculation test. But now, although speaking is not included in the exam because it is hard to manipulate, the weight of listening is up to 20% of the whole test paper.

Also, teachers will not train students' skills separately because the matriculation test pays more attention to the integrated skills such as the test task of continuation writing based on reading. Reading and writing should be combined under the influence of this question type. Writing will be upgraded to a more important position with the weight increasing from 25 to 40. And the difficulty level is highly improved, so teachers will give more emphasis on writing training.

With the deletion of multiple choice and the addition of Filling in the blanks with words, grammar teaching mode will be changed. Teachers will help students to learn grammar knowledge and how to use certain grammar points in texts and contexts, not like before, that is, teaching grammar knowledge mechanically and care about whether students can memorize them or whether they can really use them in actual contexts.

Another impact is that middle school teaching will lay more emphasis on textual ability and pragmatic ability of English because the 2014 Reform holds them in a very important position. The underlying reason is that the communicative teaching methodology and cognitivism in education have become the mainstream methodology. Modern testing and assessment will not ignore this strong influence. Various kinds of question type designs are affected, for example, when exam writers design the test paper, they tend to choose authentic language material, sometimes, they choose material from various corpora. In this way, they can guarantee the test points are not isolated from actual uses. Therefore, students should adjust their study methods. Don't just memorize the language points in isolation. Read and study authentic language material. Acquire and improve their language competence. Only by doing this can they cope with shifting events by sticking to a fundamental principle.

## CONCLUSION

The history of language assessment is the history of changes of the question types. In a certain period, the adoption of a certain question type is always backed by the theories which are popular and fashionable during that period, reflecting people's understanding and cognition of language, language learning, psychometrics and educational psychology at that time. The deepened understanding usually leads to the changes of the question types. For example, under the influence of behaviorism and structural linguistics, separate multiple-choice grammar became one of the mainstream question types (Bloomfield 1933; Lado 1961). Afterwards, with the unitary competence hypothesis being proposed, comprehensive question types, such as cloze, dictation, etc., became widespread (Oller 1979). The changes of question types will naturally bring the changes of backwash effects, with time as the catalyst in this process. All relevant factors fermenting over time force out the transformation of question types. Some question types become passivated because of long-time using, or may be eliminated because of the better choices. The history of the matriculation English test in China developed just along this line. By reviewing the history, analyzing the present and forecasting the future, we can have a better understanding of the matriculation test. This history of matriculation test can also showcase the Chinese social development and academic studies in education over these eyes. We hope we opened a window to the outer world so that you can know China better. We have made great progress while making serious mistakes at the same time, but we hope you can learn something from these painful experience.

## REFERENCES

- Baum, W. M. 1994. *Understanding behaviorism: science, behavior, and culture*. New York, NY: Harper Collins College Publishers.
- Bloomfield, L. 1933. *Language*. New York: Holt.
- Cheng, L.Y. & Curtis, A. 2010. *English Language Assessment and the Chinese Learner*. New York, NY: Routledge.
- Chomsky, N & Skinner, B.F. 1959. *A Review of B.F. Skinner's Verbal Behavior*. *Language*, 35 (35): 26-58.
- Dole, J.A., Duff, G.G., Roehler, L.R. & Pearson, P.D. 1991. Moving from the old to the new: research on reading comprehension instruction. *Review of Educational Research*, (61): 239-264.
- Friesen, N. 2005. *Mind and Machine: Ethical and Epistemological Implications for Research*. Canada: Thompson Rivers University, B.C.
- Kirkland, M. & M. Saunders. 1991. Maximizing student performance in summary writing: managing cognitive load. *TESOL Quarterly*, (1): 105-121.
- Lado, R. 1961. *Language Testing: The Construction and Use of Foreign Language Tests*. London: Longman.
- Lilienfeld, S., Lynn, S. J., Namy, L. & Woolf, N. 2010. *Psychology: A Framework for Everyday Thinking*. Pearson.

- Liu, Juesheng. 2005. The history and status quo of Matriculation English Testing in our country. *Shanghai Educational Science and Research*. (3): 40-43.
- Madden, Gregory J., ed. 2013. *APA Handbook of Behavior Analysis*. Magination Press.
- Oller, J. 1979. *Language Tests at Schools*. London: Longman.
- Piaget, J. 1967. *Logique et connaissance scientifique*. Paris: Encyclopédie de la Pléiade.
- Qi, Luxia. 2004. The washback effect of NMET. *Foreign Languages Teaching and Research* (5) : 357-363.
- Qi, Luxia. 2007. Examining the intended and actual washback of the proofreading subtest in the National Matriculation English Test. *Curriculum, Teaching Material and Method* 27(10):43-50.
- Robitaille, J. & Connelly, R. 2002. *Writer's Resources: From Paragraph to Essay*. Boston: Heinle.
- Shao, Shuiping & Hang, Yuanzhen. 2006. The status quo, difficulties and way out of NNMET. *Foreign Language Teaching in Schools* (1):2-6.
- Skinner, B.F. 1984. The operational analysis of psychological terms. *Behavioral and Brain Sciences*,7 (4): pp. 547-81.
- Wang, C. & M. Wang. 2014. Effect of alignment on L2 written production. *Applied Linguistics*.
- Wang, Chuming, & Qi, Luxia. 2013. A Study of the continuation task as a proficiency test component. *Foreign Language Teaching and Research* 45(5): 707-718.
- Wang, Chuming. 2012. Continuation task---An effective way to improve foreign language learning efficiency. *Foreign Language World* (5):1-7.
- Wang, Min & Wang, Chuming. 2014. The effects of alignment in continuation tasks. *Modern Foreign Languages* 37(4), 501-512.
- Weigle, S. C. 2004. Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, (9), pp. 27-55.
- Wu, Pengsen. 2013. A new thought on the reform of national matriculation test. *Exploration and Free Views* (4): 70-76.
- Xiao, Yunnan & Luo, Xiaoying. 2002. Reforms of English Testing. *Foreign Languages Teaching and Research* (4): pp. 294-299.
- Xu, Liang & Zou Shen 2013. Study on Assessment of Grammar in NMET (Shanghai paper). *Computer-assisted Foreign Language Education* (154),11-16.
- Yu, Shulin & Huang, Jianbin. 2009. Reflection on the test type of error correction in NMET. *Education Measurement and Assessment* (7), 53-55.