

A STUDY OF WRITING TASKS IN THE NMET 2014

Lin Dunlai

School of Foreign Languages and Literature; Collaborative Innovation Center of National Assessment and Improvement of Basic Education Quality, Beijing Normal University, China
E-mail: lindunlai@bnu.edu.cn

APA Citation: Lin, D. (2014). A Study Of Writing Tasks in The NMET 2014. *English Review: Journal of English Education*, 2(2), 145-157

Received: 01-04-2014

Accepted: 27-04-2013

Published: 01-06-2014

Abstract : The National Matriculation English Test (NMET) is a high-stakes test compulsory for those aspiring to entering higher education institutions in China. The test is designed by the National Education Examinations Authority (NEEA) and some local examination authorities. The final section of this test is generally a writing task. In this paper, the author examines the 21 writing tasks that were used in the NMET 2014. Using content analysis, and *a priori* validation components of Weir's (2005) socio-cognitive validation framework, namely test taker characteristics, cognitive validity and context validity, the paper finds out the merits and demerits of these 21 writing tasks. It is concluded that these writing tasks are generally well designed and are valid in helping select talents. But more empirical research must be done to investigate how the writing tasks function for students when taking these writing tests.

Keywords: NMET, socio-cognitive validation framework, cognitive validity, context validity, China

INTRODUCTION

Examinations play a crucial part in Chinese people's life. For Chinese students to become successful in schools or even in their life, they have to perform very well in numerous tests, especially the University Entrance Examination to Higher Education (UEEHE). The UEEHE is the largest examination system in China, which is administered annually between June 6th to 8th for most provinces, and into 9th for a few provinces. Its purpose is for general higher educational institutions to grant admission to those talented for higher education nationwide (Liu, 2010). According to Chinese Ministry of Education (MoE), the candidature for 2014 was 9.39 million.

The University Entrance Examination is administered by the National Education Examinations Authority (NEEA), which is an

institution directly under the supervision of MoE of P. R. China, and local institutions (the provincial admissions office). It normally includes compulsory subjects Chinese, Mathematics and English, and other subjects which depends on what type of university the candidate is applying for. For 28 out of 31 provinces, those other subjects like chemistry, physics, politics and geography are integrated into two tests of science or liberal arts.

The National Matriculation English Test (NMET) was designed in 1991 (Liu, 2010:38). Its predecessor, the Matriculation English Test (MET), was replaced by it in 1996 in accordance with the educational reform (see Li, 1990 for the MET; see Cheng & Qi, 2006 for an overview of the NMET). In 2014, three forms of NMET were in application. The first and second one were based on the National English Curriculum Standards

Table 1. Overview of the NMET Based on the National English Curriculum Standards (MoE, 2001)

Section	Part	Input	Lang. of rubric	Task focus	Item types	Num. of items	Raw score
I. Listening comprehension	A	5 short dialogues (heard once)	Chinese	Simple factual information	MC (3 options)	5	7.5
	B	5 dialogues & monologues (heard twice)	Chinese	Simple factual information	MC (3 options)	15	22.5
II. Reading comprehension	A	4 texts (over 900 in total)	Chinese	General and detailed information	MC (4 options)	15	45 (30)
	B	1 text (about 300 words)	Chinese	Relationship between sentences and paragraphs	Gap filling (7 options)	5	15 (10)
III. Use of English	A	1 text (about 200 words)	Chinese	Vocabulary and comprehension	MC cloze (4 options)	20	40 (30)
	B	1 text (about 150 words)	Chinese	Grammar use in context	Gap-filling (no more than three words)	10	15
IV. Writing	A	1 text (100 words)	Chinese	Error correction	Error correction	10	10
	B	Prompt (in Chinese)	Chinese	Writing a short text	Guided writing	1	25
						80+1	(150)

(NECS) (MoE, 2001). The two differ in the weighting of scores in reading comprehension and use of English sections (see table 1 for an overview of the NMET based on NECS). These two tests are informed by Lyle Bachman's 1990 model of communicative competence. The first test was used in Henan, Hebei and Shanxi, while the second test was used in Qinghai, Xizang, Gansu, Guizhou, Inner Mongolia, Xinjiang, Ningxia, Jilin, Heilongjiang and Yunnan in 2014. The third test was based on the Teaching Curriculum of English with a listening section. This test, however, is expected to die out, and was used only in Guangxi Zhuang autonomous region this year. Besides these tests, since 2004, several provinces or municipalities have been required to develop their own matriculation tests. As of 2014, 17 provinces and municipalities are developing their own test. These provinces or municipalities are Beijing, Tianjin, Shanghai, Chongqing, Liaoning,

Jiangsu, Zhejiang, Anhui, Fujian, Guangdong, Jiangxi, Shandong, Hunan, Hubei, Sichuan, Shaanxi, Hainan. Please mind that not all these provinces or municipalities developed all the tests of all subjects. They may file a request to use some of the tests developed by the NEEA. Every year, months before the UEEHE is administered, Examination syllabus for admission tests to institutions of higher education is issued to guarantee all teachers and candidates are well-informed of what is tested in the UEEHE.

A look through the past papers since 1978 showed that the writing task made its debut in 1989, and it has stayed there ever since then. The direct measures of writing assessment, which assess a student's ability to communicate through the written mode based on the actual production of written texts, is conducive to the validity of the test. The NEEA emphasizes that all tests should test what NECS requires and what is taught

in schools. In light of this, the writing task should reflect what NECS Band 8 says about what students should be able to do to achieve the proficiency required. The NECS Band 8 requires that students should be able to: write a coherent and well-structured essay to narrate or express their views or attitudes; write a summary of a text; write an essay or report on information about written texts or graphs. It also stipulates that students should have an awareness of rhetorical pattern and achieve fluency in writing.

Hyland (2002) argues that the design of good writing tasks involves four basic elements, namely, rubric (the instructions), prompt (the task), expected response (what the task intends test takers to do), post-task evaluation (assessing the effectiveness of the writing task). In China, many research papers have been published about the NEMT writing tasks. For instance, Gao and Gu (2008) examined the writing tasks set by local examination authorities from 2004 to 2007, and found problems with rubrics and communicative needs of the writing tasks. Other cross-year studies include Dong et al. (2011), Feng and Gao (2012). Zhang (2013) did a concurrent study of the NMET 2014 writing tasks. He examined the tasks from content and test design, and gave some suggestions to writing task development and classroom teaching of writing. Other similar studies are Gu and Gao (2007), Ding (2012). But we have to point out that few studies so far have used a well-established validation framework to investigate the NMET writing tasks.

The objective of this paper is to first provide a profile of what is tested in the 2014 NMET writing section across the country, and then based on Weir's (2005) socio-cognitive approach to test validation, especially test taker characteristics, context validity and cognitive validity, the paper examines the quality of these writing tasks and

give suggestion where improvements could be made. In dealing with these aspect, we only focus on what Weigle (2002: 77-78) called the design stage and operationalization stage.

METHOD

As mentioned above, administered by NEEA, the NMET is a compulsory subject that every candidate aspiring to entering higher education institutions should take. NEEA is not responsible for developing all tests that are administered across the country. In 2014, 17 provinces and municipalities developed their own tests, including the NMET. There are two versions of tests developed by the NEEA, based on the NECS. The third test based on the Teaching Curriculum of English, used in Guangxi, shared the same writing task as one of the tests based on the NECS. Besides, the NMET of Beijing, Guangdong, Chongqing includes two writing tasks each. Hainan used the second test developed by the NEEA. Put together, there are 21 writing tasks in the NMET of 2014. These 21 tasks are the samples for analysis for this paper. Table 2 gives us a profile of general information about these 21 writing tasks.

According to Coombe et al. (2007: 73), writing prompts can be divided into three types: base prompts, framed prompts and text-based prompts. As is shown above, framed prompts are most frequently used in the NMET 2014. Base prompts only appeared in Fujian paper, Hubei paper, Shandong paper, Chongqing paper Task1. Text-based prompts appeared in Jiangsu paper and Guangdong paper.

How well does these writing tasks function in such a high-stakes national test? In order to answer this question, we should seek a test validation framework to analyze the above 21 writing tasks. We believe that the socio-cognitive framework proposed by Weir (2005) and refined later through the experience of

applying it to operational Cambridge ESOL tests to be both theoretically sound and extremely useful. For the current paper's purpose, we find test taker characteristics, cognitive validity (originally named theory-based validity) and context validity proposed by Weir (2005) and exemplified in Shaw and Weir (2007) to be relevant. This part of the socio-cognitive framework is presented in Figure 1.

As shown in Figure 1, *a priori* (before-the-test event) validation components, as named by Shaw and Weir (2007: 3) include test taker

characteristics, cognitive validity and context validity. The three crucial research questions that we can pose are:

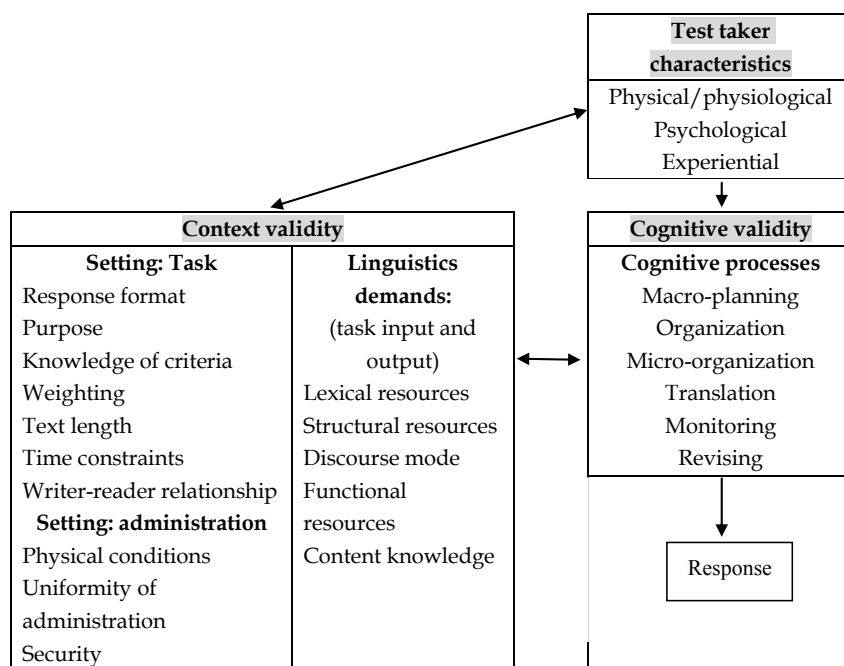
- How are the physical/physiological, psychological and experiential characteristics of candidates catered for by the NMET writing tasks? (test taker characteristics)
- Are the cognitive processes required to complete the test tasks appropriate? (cognitive validity)
- Are the characteristics of the test tasks and their administration appropriate and fair to the candidates who are taking them? (context validity)

Table 2 A profile of 21 writing tasks in NMET 2014

Name	Number of items	Raw score	Length	Topic
NECS1	1	25	~100*	A letter of inquiry about study in the UK
NECS2	1	25	~100	An essay about what life will be like in ten years
Guangxi	2	15	>50	A letter of apology and further arrangement of travel
Beijing			>60	An essay about how a problem was solved
Tianjin	1	25	>100	An email to an exchange students about local life
Shanghai	1	25	120-150	An email giving suggestion and explanation
Chongqing	2	15	>60	An essay expressing views about a saying
		20	>80	A post about the best gift you received
Liaoning	1	25	~100	A written announcement about an activity on campus
Jiangsu	1	25	~150	A summary and an argumentation about word borrowing
Zhejiang	1	30	100-120	A letter to a librarian to recommend books
Anhui	1	25	~120	A response letter to a worried person
Fujian	1	25	~120	An essay of views about a quotation about success
Guangdong	2	15	n/a	A report of five sentences to a local English paper
		25	~150	A summary and students' views about donation
Jiangxi	1	25	~120	A speech about what to learn in senior high school
Shandong	1	30	120-150	An essay about a proverb (choose one from two proverbs)
Hunan	1	25	>120	An introduction to your innovation for a daily item
Hubei	1	30	>120	An essay about "small things make a difference"
Sichuan	1	35	~120	A letter to a pen pal about college entrance exam
Shaanxi	1	30	>100	A welcoming speech to a foreign delegation

* ~stands for "approximately", >stands for "more than"

Figure 1. *A priori* validation components for conceptualizing writing test performance (adapted from Weir, 2005:47)



RESULTS AND DISCUSSION

In this part, we will examine the quality of the 21 writing tasks that appeared in the NMET 2014. We will follow the three research questions as we discuss about the quality. For each question, a brief introduction to the contents of the validation components will be put forward, followed by a discussion about the 21 writing tasks.

Test taker characteristics

Shaw and Weir (2007:17) argue that it is the test taker, rather than the test task, that is at the heart of the assessment event. While test takers' success in the test depends primarily on an individual's ability in the intended construct, there are many other variables which are likely to impact on test takers' performance. One of these variables is test takers' personal characteristics, which includes factors such as age, interests, experience, knowledge and motivation. Bachman (1990) argues

that test taker characteristics are either systematic in the sense that they influence test performance continually, and unsystematic in the sense that they are more difficult to predict. The systematic attributes referred to by Bachman (1990:164) include cognitive style, content knowledge, and gender. O'Sullivan (2000) listed the following test taker characteristics, as presented in Table 3.

As for physical/physiological aspect, test takers for the NMET are mostly of the same age, around 17 to 19 years old. These school-leaving students were born around 1995, when multimedia, such as computers and the internet, are becoming an important part of their life. This was reflected in some writing tasks in the NMET of 2014. For instance, the two writing tasks in Chongqing paper were about phone and posting information on the internet. The first task use a popular saying "My friend falls, I laugh; my phone falls, I cry" to

Table 3. Test taker characteristics (based on O’Sullivan, 2000)

Physical/physiological	Psychological	Experiential
Short term ailments	Personality	Education
Longer term disabilities	Memory	Examination preparedness
Age	Cognitive style	Examination experience
Sex	Affective schemata	Communication experience
	Concentration	Target language
	Motivation	Country residence
	Emotional state	

arouse consensus of a generation of nomophobia. The Jiangsu paper gave a text about the intrusion of English words like “pk” (meaning to compete), “out” (meaning ill-informed) into the latest dictionary of contemporary Chinese, causing a heated discussion. Students at this age are most active in using these new words, and they have their views about it.

As for psychological aspect, test taker’s interest or motivation may influence the way a task is dealt with (Berry, 2004). Test developers should make efforts to make test events as positive as possible. There was a bad example. For such a high-stakes test, students and parents alike are suffering a lot of anxiety. But the Sichuan paper asked test takers to write a letter to a pen pal about college entrance exam. It may arouse more anxiety and cause construct irrelevant variance. To cater for the candidates with different cognitive style, NECS1 paper, Liaoning paper, Guangdong paper task I used a diagram, and Beijing paper Task 2 used pictures.

As for experiential aspect, test takers should be informed of the demands of the writing tasks. This is annually explained in *Examination Syllabus for Admission Tests to Institutions of Higher Education* (NEEA, 2013). An examination of the 21 writing tasks show that all these writing tasks should be familiar to test takers. Experiential characteristics also includes students’ past learning experiences. The NMET 2014 was paper-based and students spent a long period of time revising,

so there is confidence that they were familiar with what was tested in the writing tasks. But we have to bear in mind that China is such a large country, even within a province, the situations for students could be dramatically different. Students in cities may have communication experience with English speaking people, while students in the countryside might not have the chance to come across a foreigner. Cheng and Qi (2006) provide a good example about a writing task requiring the candidates to describe an experience of violating traffic regulations, thus biasing against the candidates from the countryside. An examination of these 21 writing tasks shows no obvious bias against the candidates from the countryside in the NMET 2014.

Cognitive validity

The cognitive validity of a writing task is a measure of how closely it represents the cognitive processing involved in writing contexts beyond the test itself, i.e. in performing the task in real life (Shaw & Weir, 2007: 34). Scholars have proposed many frameworks of writing, especially for L1 writing (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987; Kellogg, 1994, 1996; Field, 2004; Eysenck & Keane, 2005). Grabe and Kaplan’s (1996) work is part of a limited literature available for cognitive processing in L2, but as Shaw and Weir (2007) points out their model does not distinguish adequately the resources stored in long-term memory from the operations of short-term memory. Field

(2004:329-331) provides an account of information processing which shows the operations a writer performs when engaged in the writing process. It was adopted by the above Weir (2005) model, and we will use this model here, and as we use content analysis to analyze the writing topics, we focus on macro-planning, organization, translation and monitoring and revising.

macro-planning: *gathering of ideas and identification of major constraints (genre, readership, goals)* (Shaw & Weir, 2007:38)

Scardamalia and Bereiter (1987) argue that different processing models exist at different developmental stages of writing and describe two major strategies, *knowledge telling* and *knowledge transformation*. In knowledge telling, the writer does not plan much and only focuses on generating content from within remembered existing resources in line with task, topic or genre. Knowledge transforming entails a heightened awareness of problems as and when they arise. The writer considers the complexities of a task and analyze and solve problems of task achievement in terms of content, audience, register and set goals (Hyland, 2002:28).

An examination of the 21 writing tasks show that the planning, processing, retrieval and evaluation demands upon the NMET candidates are generally simple and largely involve knowledge telling. Take NECS2 which was most widely used for an example. The writing task requires the candidates to write an essay about "My life in ten years" for an English newspaper. They were required to follow three points, namely, family life, work, entertainment. The first sentence is given as "I often imagine what my life will be like in the future". Macro-planning of this task may entail: an assessment of the context (the need for an essay to the local English newspaper for publication);

appreciation and understanding of the three points in the input text (family life, work, entertainment); a preliminary representation of the writing outcome (a 100-word essay); an evaluation of the potential problems in understanding the task (understanding input text, coverage of content bullet points, appropriate essay for publication); an initial activation of certain aspects of the genre; perhaps also strategic considerations-avoiding what they cannot express. The other framed prompts are similar to NECS2. However, those base prompts and text-based prompts required more planning and involved knowledge transformation. One example is from Hubei paper, which asked the candidates to write an essay based on the following statement "Small things make a big difference. The small things we do can make us a responsible member of the society." Macro-planning for this task may entail: an assessment of the context (an essay about small things, should include specific events of the candidate); an initial choice of the genre, whether it is to narrate or to argue; a preliminary representation of the writing outcome (a 120-word essay); an evaluation of the potential problems in undertaking the task (understanding input text, choosing an appropriate event to support the idea); an initial activation of certain aspects of the genre: how narrative is structured.

organization: *ordering the ideas; identifying relationships between them; determining which are central to the goals of the text and which are of secondary importance*(Shaw & Weir, 2007:38)

Skilled writers plan the writing in relation to the overall text and points within the writing at the initial stage of the writing process, while unskilled L2 writers experience a heavy cognitive load in encoding their thoughts in linguistic form so that they normally can't manage

to pay attention to organizing the idea. In the NMET 2014, most prompts are framed prompts, so the demands on the test takers to organize their thoughts is low. For instance, in NECS1, a letter of inquiry to study in the UK in summer holiday is required. The points that need to be included were given: Start date? Class size? Hours per week? How much? Types of accommodation? The candidates only needed to follow these points to come up with the letter. Not much attention should be paid to organization. However, with base prompts, the story could be different. In Chongqing1, the candidates are asked to make comments on "My friend falls, I laugh; my phone falls, I cry." No points were advised in the prompt, and the candidates should plan for the organization by themselves.

translation: *propositional content previously held in abstract form is converted to linguistic form* (Shaw & Weir, 2007:39)

At this stage, the propositions in the candidates' mind is transformed into linguistic forms. This is basically automatic and is generally not subject to direct investigation. However, we can infer from the marking schemes provided by some writing tasks to know the requirements. Again we use NECS1 for example. The marking scheme shows that for those skilled writers, they should use various grammatical structures and a wide range of vocabulary. They may make a few mistakes, but these mistakes result from a good attempt to use advanced grammatical structures and vocabulary. They should also need to use cohesive devices to make the text hinge together. The marking scheme also shows that for those who are not able to achieve the task, they use very simple grammatical structure and very limited vocabulary. They make many mistakes in grammar and vocabulary, making it impossible to communicate and they

are not able to use cohesive devices. But we need to keep in mind that even for skilled L2 writer, for students at this level of proficiency, they are encouraged to be ambitious in their use of language, thus good attempts with minor mistakes in language generally do not lead to punishment.

d. monitoring and revising: *a basic level of checking mechanical mistakes, and a higher level of checking task fulfillment, and make some modifications accordingly.*

In test preparation period, English teacher would make efforts to ask the candidates to check their writing before handing in. For unskilled writers, they normally focus on checking spelling, punctuation and syntax. For more skilled writers, they may consider whether their writing is fit for the situation, and may go back to those aspects of dissatisfaction, and make some adjustments.

Context validity

Context validity relates to the linguistic and content demands that must be met for successful task realization (Shaw & Weir, 2007:63). Weir (1993) argues that the sample of communicative language ability selected for a test should be "as representative as possible" and the test tasks should be selected in accordance with "the general descriptive parameters of the intended target situation particularly with regard to the skills necessary for successful participation in that situation". Based on the model proposed in Figure 1, this section will examine the NMET 2014 writing tasks from setting: task, setting: administration and linguistic demands: task input and output.

Setting: task

Rubrics and prompts. In writing tasks, test developers must make sure that any production demands on the test

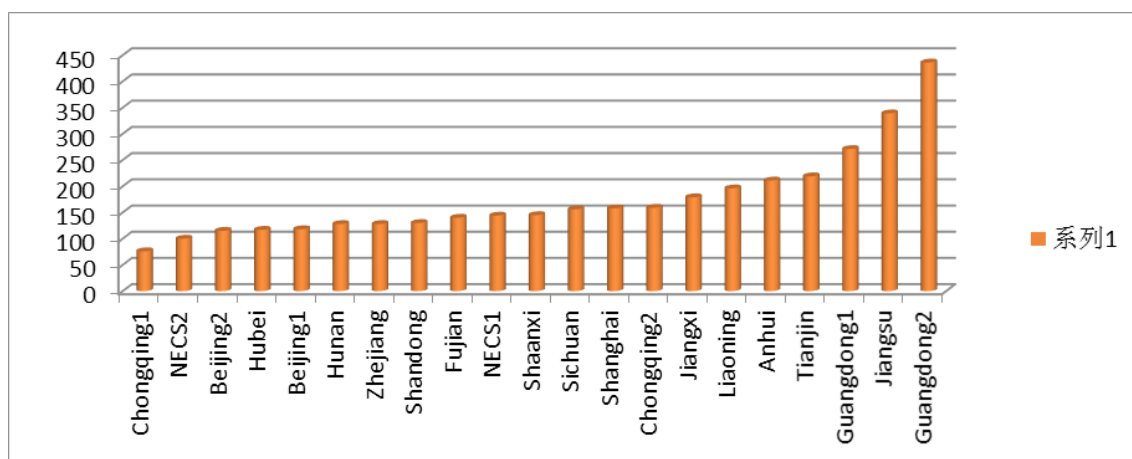
takers are made clear. Bachman (1990: 118) suggests three characteristics of task rubrics: test organization, time allocation and instructions. Later, Bachman and Palmer (1996: 121) offer three essential guidelines for instructions. They argue that instructions should be a) simple enough for test takers to understand; b) short enough so as not to take up too much of the test administration time; c) sufficiently detailed for test takers to know exactly what is expected of them.

From Figure 2, we can see that most rubrics plus prompts do not exceed 150 words. The candidates using NECS2 only needed to read 100 words in Chinese, thus the reading load was kept to minimum. The two longest were Jiangsu paper and Guangdong paper. These two papers used text-based prompts, with the length of the texts being 168 and 208 respectively.

Authenticity. The authenticity of a test is of paramount importance for the validity of it. Authenticity must necessarily be seen as a relative, rather than an absolute, quality, since the context of a testing situation is greatly different from that of typical real-life situations (Elliott & Wilson, 2013:154). Lewkowicz (2000) states that authenticity can be a function of

both the input and the output, and it is possible for a test to have a high level of authenticity in one but not the other. Bachman (1991:690-691) proposed two notions of authenticity. One is *situational authenticity*, which means “the perceived relevance of the test method characteristic to the features of a specific target language use situation”. The other is *interactional authenticity*, which means “the interaction between the test taker and the test task”. The interactional authenticity is later renamed as interactiveness in the test usefulness framework proposed by Bachman and Palmer (1996). An examination of the 21 writing tasks show that four tasks violates situational authenticity. The first one appeared in Zhejiang paper, which asked the test takers to write a letter to the school librarian Mr. Wang (a typical Chinese) to recommend two categories of new books. This simply could not happen in Chinese campus where it is rare to have an English speaking person to be a school librarian. Besides, when it comes to recommendation, it is more efficient to write in Chinese. Another is Anhui paper which asked the candidates to write a response letter to the worried person. Although it is a column in the campus English newspaper, rarely do

Figure 2. Word count of rubrics and prompts of the NMET 2014



Chinese students express their concerns through the campus newspaper. The third is Liaoning paper, which asked the candidates to write an announcement about innovation on campus. As this activity has nothing to do with English, it is not necessary for students to write an English announcement about it. The fourth is Hunan paper, which asked the candidates to write about their innovation for a daily item with the title "my magic ____". As innovation would necessarily involve something creative, and it is very difficult for the candidates with this level of language proficiency to describe it well and use it for a competition.

Response format. Alderson et al. (1995) suggest that a test should include a range of response formats in order to ensure that all candidates will have an opportunity to perform at their best. In Table 1, we know that the writing section normally have two items, one is error correction and the other is a writing task. When we examine the NMET 2014, we see that a few provinces and municipalities such as Beijing, Chongqing and Guangzhou used two tasks. And in Shandong paper, two proverbs were offered, "A friend in need is a friend indeed", or "Where there is a will, there is a way." The candidates are required to choose one of the two proverbs and make comments.

Purpose. The rubric must give the candidates clear and precise information about the purpose for completing the writing task. Weir (2005) argues that there is a close relationship between the choices we make in relation to purpose and the processing that results in task completion. In this regard, all the 21 writing tasks are doing a good job, as they state very clearly what is expected from the candidates, either it is a letter of inquiry, a welcoming speech or an essay to be published.

Knowledge of criteria. Weir (2005) also

emphasized the importance of letting candidates know which criteria are used in the marking. Only two papers state explicitly the marking criteria on the test paper. One is Jiangsu paper and the other is Guangdong paper. For instance, Jiangsu paper states that scores will be awarded for content completeness, language, discourse coherence and text length.

Weighting. Weighting should be stated clearly on the test paper so that candidates will arrange their time in the testing process. All papers are doing very well in this regard except the Shanghai paper. The test paper only shows that the second section with a translation test and a writing task is worth 47 points without indicating clearly the weighting for the writing task.

Text length. As is shown in Table 2, all writing tasks show the requirement of text length except the Guangdong paper Task1. But the wording for text length varies across different writing tasks. The words "approximately" and "no less than" are frequently used. Zhejiang, Shanghai and Shandong papers give a range of the text length. We think that the word "approximately" is more confusing, as there is not an exact word number to follow. Giving a range of text length requires test takers to have a better planning for their writing.

Time constraints No paper indicates time constraint for the writing task. The whole test paper is supposed to be finished within 2 hours. It gives students freedom to make decisions on how much time to spend on each section. This also makes great demands on the candidates' test taking strategies.

Writer-reader relationship Writing is a kind of social interaction, and there should always be an audience for a certain piece of writing. In the 21 writing tasks, we find that the distance between the writer and the audience are different, thus making the tasks different

in difficulty. 13 of the tasks use an open audience, and the candidates need to consider how their writing can convey their ideas properly. The rest of the tasks involved a specific person who is not very familiar to the candidates, thus calling for proper formalities.

Setting: administration

Security is very tight for the administration of the national college entrance examinations. Test papers are confidential until they are put into real use. In the writing tasks, a sentence suggesting “no information should be stated about your name and school name” often appears in the rubrics. Out of the 21 writing tasks, 10 tasks stated requirement. An imaginary name is used where necessary. 7 test papers use *Li Hua* as an imaginary name for the test candidate to guard against possible release of the candidates’ real names. The other names are “*Li Jin*”, “*Jamie*”, “*Li Yue*”, “*Li Xia*”, “*Joe*”, “*Cheng Fei*”. All these measures are taken to guarantee the validity of the test.

Linguistic demands: task input and task output

As for linguistic demands, including lexical resources, structural resources, discourse mode, functional resources and content knowledge, it is clearly stated in the syllabus for the NEMT (NEEA, 2013) in five attachments to the test specification.

An examination of the 21 writing tasks shows that in terms of text types, genres and topics, the tasks were developed strictly based on the syllabus. Any new words that appeared in the writing tasks were indicated in the instruction, such as “accommodation”, “innovation”, “literary book”, “quotation”.

The language of input, including rubrics and prompts, differs in the writing tasks. NECS1, Fujian paper,

Hubei paper, Shandong paper and Chongqing paper Task1 used Chinese rubrics and English prompts. Shanghai paper, and Hunan paper used English rubrics and Chinese prompts. NECS1, Liaoning paper, Anhui paper, Jiangxi paper, Shaanxi paper, Zhejiang paper, Sichuan paper, Beijing paper Task 1 and Chongqing paper Task2 and Guangdong paper Task1 used only Chinese input. Jiangsu paper and Guangdong paper Task2 used English text plus Chinese input. Liaoning paper and Beijing paper Task2 used Chinese input with pictures or graphs.

CONCLUSION

Adopting *a priori* validation components from Weir’s socio-cognitive framework for test validation, we have found that the 21 writing tasks are generally valid in selecting university entrants. The analysis from the perspective of test taker characteristics shows that the candidates’ cognitive style, content knowledge and their age have been catered for, though a few tasks may arouse anxiety. The analysis from the perspective of cognitive validity shows that framed prompts normally require knowledge telling while base prompts and text-based prompts are more likely to entail knowledge transformation. The analysis from the perspective of context validity shows that test developers have made a lot of efforts to make the prompts briefly and clearly. Some tasks suffer from situational unauthenticity. A few writing tasks give candidates freedom to make the choice of topics. As regards audiences, the 21 writing tasks have great difference in light of the imaginary writer’s familiarity with the audiences, thus leading to different difficulty. Security in test administration is guaranteed to avoid construct irrelevant variance. Task input and output are strictly based on NECS and within the

candidates' ability to tackle. However, the current paper only adopts content analysis, and nothing has been done as to how the candidates process these writing tasks. Future research should focus on this to find out more evidence to validate the writing section in the NMET.

REFERENCES

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University 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), 671-704.
- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bereiter, C. & Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berry, V. (2004). *A Study of the Interaction between Individual Personality Differences and Oral Performance Test Facets*. Unpublished PhD Dissertation, the University of London.
- Cheng, L. Y. & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly*, 3(1), 53-70.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A Practical Guide to Assessing English Language Learners*. Ann Arbor: The University of Michigan Press.
- Ding, Y. (2012). A concurrent study of the writing tasks of the NMET 2011. *Educational Measurement and Assessment*, 3, 55-59.
- Elliott, M. & Wilson, J. (2013). Context validity. In A. Geranpayeh & L. Taylor (Eds.). *Examining Listening*. Cambridge: Cambridge University Press, 152-241.
- Eysenck, M. & Keane, M. (2005). *Cognitive Psychology* (5th Ed.). Hove: Psychology Press.
- Field, J. (2004). *Psycholinguistics: the Key Concepts*. London: Routledge.
- Gao, X. & Gu, X. (2008). Characteristics and problems of the NMET writing tasks set by local examination authorities. *Foreign Language Teaching in Schools*, 3, 18-22.
- Gu, X. & Gao, X. (2007). A concurrent study of the writing tasks of the NMET 2007. *China Examinations*, 6, 28-33+36.
- Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.). *The Science of Writing*. Mahwah, NJ: Lawrence Erlbaum Associates, 1-28.
- Hyland, K. (2002). *Teaching and Researching Writing*. London: Longman.
- Kellogg, R.T. (1994). *The Psychology of Writing*. New York: Oxford University Press.
- Kellogg, R.T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.). *The Science of Writing*. Mahwah, NJ: Lawrence Erlbaum Associates, 57-72.
- Lewkowicz, J. (2000). Authenticity in language testing: some outstanding questions. *Language Testing*, 17(1), 43-64.
- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11, 393-404.
- Liu, Q. (2010). The National Education Examinations Authority and its English Language Tests. In L. Cheng & A. Curtis (Eds.). (2010). *English Language Assessment and the Chinese Learner*. London: Routledge, 29-43.
- NEEA. (2013). *Examination syllabus for university entrance examination to higher education for 2014*. Beijing, China: Higher Education Press.
- O'Sullivan, B. (2000). *Towards a model of performance in oral language testing*. Unpublished PhD Dissertation, University of Reading.
- Scardamalia, M. & Bereiter, C. (1987). Knowledge telling and knowledge transformation in written composition. In S. Rosenberg (Ed.). *Advances in Applied Linguistics, Vol. 2, Reading, Writing and Language Learning*. Cambridge: Cambridge University Press, 142-175.
- Shaw, S. D. & Weir, C. J. (2007). *Examining writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.