Test Proposal for a Second Language Test: A Computer-Based Test (CBT) of Writing

Sookhee Jeong

Colorado State University

Abstract

The trend of English language tests in Korea focuses on measuring students' communicative competence with a computer-based test (CBT). The Korean government is implementing a CBT English test, called the National English Aptitude Test (NEAT). The purpose of this test proposal is to measure students' integrative writing ability; the test is designed for a CBT. This proposal provides all the information about a CBT achievement writing performance test, from the purpose of the test to discussion with critiques of item performance. This test proposal would present an ideal NEAT-typed integrated writing test tasks with high reliability and validity.

Keywords: computer-based test (CBT), National English Aptitude Test (NEAT), integrative writing test

Test Proposal for a Second Language Test: A Computer-Based Test (CBT) of Writing

Introduction

Importance of Investigating the Language Area and Test Technique

This project focuses on developing an integrative writing test for high school students in Korea that can later be administrated via computers. It is assumed that one of the most difficult skills for English as a Foreign Language (EFL) learners and international students who are studying in English speaking countries is writing. That is because the writing convention is different from languages. In other words, English has a different writing convention from Korean.

According to Grabe and Kaplan (1989), writing is affected by cultural differences including "rhetorical patterns of organization, composing conventions, cohesion and coherence patterns, writing conventions affecting choice and frequency of text types" (as cited in Connor, 2011, p. 65). Kaplan (1966) analyzes foreign college students' writing and claims that each foreign student showed the typical logic and the language patterns based on each of their own cultures. Namely, their mother language and cultures affect on their English writing. Kaplan (1966) strongly argues that L2 learners need to know the logic of the English as well as the form of English. Therefore, it is valuable to develop a writing test while taking these different factors.

When assessing writing ability, an important decision is that writing tasks in the target language use (TLU) domains require language users to use integrative skills. They need to take notes while listening to lectures, to summarize what they hear, or to write their opinions after reading a text. Wolf & Wolf (2002) claim that a writing test should be considered with the purpose of "preparing students to meet writing requirements for any purposes they may encounter throughout their lives" (as cited in Higgins, Miller, & Megmann, 2006, p. 313).

Unfortunately, English language education in Korea has been largely oriented towards the

Korean Scholastic Aptitude Test (KSAT), which mostly focuses on testing receptive skills such as listening comprehension and reading comprehension (Joe, 2011). Therefore, school curricula across Korea are focused on listening and reading, but less attention is paid to speaking and writing. As a result, high school graduates in Korea demonstrate low speaking and writing ability compared to reading and listening ability (Joe, 2011).

The current Korean government has tried to replace the KSAT by focusing on a communicative approach to improve students' communicative competence. That is why the National English Ability Test (NEAT) is being developed and is to be implemented in 2015, which will replace the KSAT. NEAT is an Internet Based Test (IBT) designed to test four English skills: listening, reading, speaking and writing (Jeong, 2012). Because NEAT is now in the preparation stage, nobody can precisely predict an impact that the new test will have. The good point, however, is that students will have the chance to acquire all four skills of English (Joe, 2011).

Since test items assessing writing and speaking ability have not been a main interest in Korea, and NEAT is the internet-based test instead of a traditional pencil-and-paper test, it will be revolutionary in the history of English assessment in Korea. However, one of the pressing issues for administrators and teachers is creating computer-based writing and speaking tasks. Because writing tests are much less dealt with in English language education in Korea compared to speaking tests, this proposal will focus on writing test items reflecting the types of writing tasks likely to be in NEAT.

Method of Organization of the Paper

This project includes NEAT-typed integrated writing test tasks and it provides information about description of the test in terms of purpose of test, type of test, interpretation of

scores, TLU domain, construct definition, design of test, description of test tasks, item writing, editing and development rubrics. Next, the pilot test procedure is presented with participants, administration, and scoring procedures. Then, proposed test results are provided in terms of item statistics, descriptive statistics, reliability, standard error of measurement (SEM), and description of master/non-masters. Lastly, discussion is presented with critique of item performance, evaluation of test usefulness, overall estimation and personal reflection.

Description of the Test

The test is a Computer Based Test (CBT) test which takes place within the school computer lab. It is an achievement test which measures students' writing ability based on course objectives. The target population is 10th grade students at Korean public high school and their writing score is included in their final English grade.

Purpose of Test

Test purpose can be defined in regards to three aspects: inference, use, and impact (Stoynoff & Chapelle, 2005). First, in terms of inference, the test is intended to measure students' integrative writing ability. Second, the test score is used as their final performance score, which is combined with a paper-based test score for their final English score at the end of the semester. The test results are used as part of students' final English score, and their final score is used to make decisions about their college admission in addition to KSAT scores. Last, according to Bachman and Palmer (1996), impact can be considered in two ways: a micro effect on individuals and a macro effect on society and education system. The test focuses on communicative competence, so it is assumed that the test results will have an impact on students in terms of motivation, language ability and communicative competence. CBT-type test is still new to students, teachers, parents, administrator and society, and it might take some time for the

test to settle down after many trials and errors.

Type of Test

It is a summative test that will be used for assigning grades at the end of a semester. The test will be used to describe what students can and cannot perform in the given writing tasks. Furthermore, it will indicate mastery of students' learning outcomes, too.

Interpretation of Scores

The interpretation of the results will be criterion-referenced, which describes students' performance in terms of a specified domain of clearly defined learning tasks rather than describing students' performance according to relative position in some group (Miller, Linn, & Gronlund, 2008).

Target Language Use Domain

A TLU domain refers to language use that is relevant to the test takers' actual life and that can be used in specific situations beyond the test itself (Bachman & Palmer, 1996). Namely, the characteristics of TLU tasks and test tasks show how performance on a language test is related to language use in specific situations. Therefore, identifying the characteristics of the TLU domain and test task is very important because they are related to authenticity and validity of the test, which are both important characteristics of test usefulness.

The TLU domain is defined by the fact that students are studying in English speaking countries. They need to be able to summarize a text, to express their preference and to write a paragraph giving an advice in their Language Arts class. It is assumed that students are required at least to write a paragraph made up of 60-80 words as well as to write a paragraph while listening or reading a prompt. The instructor also expects students to use appropriate vocabulary, grammar and to organize the text using proper transition words. An analysis of TLU tasks and

test task shows a fairly high correspondence with the test task characteristics (Appendix A).

Construct Definition

When developing a test, we need "to specify particular components of the ability or abilities to be measured in a way that is appropriate to a particular testing situation" (Bachman & Palmer, 1996, p.116). For this writing test, the test takers are required to summarize a text, to write a short reply giving advice, and to write a paragraph to an alternative question with supporting sentences. Based on Bachman & Palmer's (1996) model of communicative language ability (1966), the specific components of language ability in the construct definition are presented in Table 1. These components will be used for the rubric when students' responses are scored.

Table 1

Components of Language Knowledge

Language Knowledge	Description
Grammatical Knowledge	-Lexis and Syntax: ability to comprehend or accurately arrange appropriate words at sentence level
Textual Knowledge	-Organization: ability to comprehend organizational development in texts or to get texts organized structurally
Functional Knowledge	-Ideational: the ability to express particular ideas; expressing one's preference; summarizing a dialogue -Instrumental: the ability to get someone to do something; giving advice
Additional Information	-Strategic competence, listening and reading ability, and sociolinguistic knowledge (register) are assumed and speaking ability is not included.

Design of Test (Table of Specifications)

The table of specifications (TOS) offers the scheme for the test in that the TOS helps when developing tasks because it includes "the content areas and the objectives we wish to measure" (Miller et al., 2008, p. 77). The TOS shows the instructional objectives and content at the same time, which are essential for designing test tasks. The TOS for this proposal (see Appendix B) reflects the scoring criteria for the rubric from the task objectives such as grammatical knowledge (vocabulary and language use), textual knowledge (organization), and functional knowledge (ideational and instrumental) as well as the topic of the task, such as giving advice to a friend, a preference about shopping style, and summarizing a dialogue about daily life.

Description of Test Tasks

As this will be a computer-based test, the specific instructions about the procedures for taking the test and scoring are provided aurally and verbally on a computer monitor. Because the CBT-type test is still new to students, the students will have had enough time to practice operating the computer system to get accustomed to it in advance. In addition, the handout about specific test instructions written in Korean and English will be distributed to each student beforehand to make the instructions more comprehensible. Students have 5 minutes for listening to directions, including time to check their headsets and computers. Students are provided with scratch paper and they should type their answers on the computer, which will be saved and sent to the teachers automatically.

The writing test is made up of three writing tasks. Each task is an integrative type, so input for each task is presented as either audio (task 2 & 3) or written text (task 1). According to the prompts, students are required to type answers in 60~80 words within the given time. In terms of the relationship between input and response, the writing tasks are non-reciprocal because students' response does not affect the form of subsequent material. The scope of

relationship is broad because students use a lot of input rather than focusing on specific details.

In the case of task 3, the response is direct because the students' response to task 3 is largely

from information provided in the input. Meanwhile, in the case of task 1 & 2, the response is

indirect because the students' response to task 1 & 2 is from information in the context or in their

own topical knowledge instead of relying on information provided in the input. The overall

features of the writing test are shown in Table 2.

Table 2

Description of Overall Writing Test

Features	Task 1	Task 2	Task 3
Item types	Restrictive Response (reading & writing)	Restrictive Response (listening & writing)	Restrictive Response (listening & writing)
Topic	Giving advice to one's friend	Writing preference about one's shopping style	Summarizing a dialogue about daily life
Administration time	15 min.	15 min.	15 min.
Word limit	60~80	60~80	60~80

Students have 45 minutes to complete 3 writing tasks, of which 15 minutes are allotted to

each task and 5 minutes for directions. Considering the fact that students usually write 60-80

words in 15 minutes in class and the regular class is 50 minutes, the time allotment would be

appropriate. If they finish each task early within the given time, they can move on to the next

task. However, once they move to another task, they are not allowed to go back to previous tasks,

which students are informed of in advance.

Item Writing & Editing, Development of Rubrics

The writing test is composed of 3 integrative tasks; one is a reading-writing task and the

others are listening-writing tasks. The reading-writing type is presented first, and the latter

follow. Rubrics are adapted from Analytic Essay Rubric for Classroom Use based on NorthStar3 Writing Rubric and ESL Composition Profile (Jacobs et al. 1981). The rubrics include grammatical knowledge, textual knowledge and functional knowledge and the rubric scale ranges from 0~4 for each criterion and 0 points are given to the students with no response at all. The total possible score is 36 points and the actual rubrics are presented in Appendix D.

Pilot Test Procedure

Participants

Test takers are seventeen-year-old public high school students in Korea and include both males and females. They have been studying English as a foreign language at school for eight years and most of them go to a private English academy after school. They have an English conversation class once a week, and reading class five days a week now and each lesson is 50 minutes. The students' English proficiency shows a wide range from beginner to advanced. Since the examinees study English for the university entrance exam, which focuses on reading and listening comprehension, they are familiar with paper-based reading and listening tests but they have never taken computer-based tests at school until now. Therefore, they have little confidence in their writing ability.

Administration

The test includes three tasks which are administered at the end of each semester in the computer lab at the school. Teachers make test items using software supplied from the ministry and all the test items are presented on computer. Students are required to type each answer on individual computers. The individual headsets are provided to students for integrative listening and writing tests. The whole test time is 45 minutes with 15 minutes assigned to each task; preparation time for checking the computer system, such as sound or keyboards, is not included

in the test time. Prior to taking the test, all students are informed about the purpose, the

procedure and scoring of the test by the teacher and the written guidelines about the test are

distributed to all the students in advance. In addition, students have a chance to take two

preliminary tests in the computer lab to get accustomed to the computer-based test format.

Scoring Procedures

Students' responses will be sent to two teachers who are in charge of grading their

responses. Students' answers will be scored by two raters who received professional training

using an analytic rubric based on grammatical knowledge (vocabulary and language use), textual

knowledge (organization), and functional knowledge (ideational and instrumental). The average

score of 2 raters will be the students' final score.

Proposed Test Results

Descriptive Statistics

Descriptive statistics describe the distribution of data and summarizes students'

performance and I will use the mean, the standard deviation (SD) and the range. First, the mean

is the most useful and widely used statistic for describing central tendency and it provides the

average of a set of scores (Miller et al., 2008, p.501). I expect I can get the average score of all

students' data through the mean. Next, the SD refers to an average distance of all scores from the

mean, so I can get information about the spread of students' scores. Finally, because the writing

test I designed is scored by two raters, the range of scores will give another indication about

variability of students' scores with regards to two raters by providing a distribution between the

highest score and the lowest score of students. It is assumed that I can get useful information

whether teachers use the full range of the scoring scale and they are strict or lenient for grading.

Reliability and Standard Error of Measurement

I will use the Pearson correlation coefficient for the inter-rater reliability to estimate reliability for consistency of judgmental scores by two raters. It is noted that well-defined agreed-on scoring rubrics and professional training of raters are essential to achieve satisfactory levels of inter-rater consistency and to prevent low inter-rater reliability derived from differences in the severity and strictness of rating from two raters. An acceptable value for criterion-referenced test would be $\geq$.60.

The SEM is used to estimate the amount of variation to be expected in assessing scores and it indicates the amount of error to be considered in interpreting students' scores (Miller et al. 2008). I will use the SEM because it is related to the reliability of scores. When the SEM is small, it indicates high reliability and it means the students' obtained score is close to their true score that is a hypothetical score free of error. Therefore, it is noted that I should increase the reliability of the test to lessen the amount of error in interpreting the students' score.

Description of Master/Non-masters

When we determine the students' mastery of the given objectives, we need to consider many factors such as the reliability, the mean and the SD. There are two kinds of charts which estimate the mastery without hassle: the agreement coefficient and kappa efficient (Subkoviak, 1988). Subkoviak provides a table to estimate the mastery based on z-score and the reliability of the test score (*r*). I will use the agreement coefficient to estimate the students' mastery that measures overall consistency. The agreement coefficient chart provides the expected portions of masters and non-masters by means of z-score and the reliability of the test score (*r*) from only a single test result. In addition, it is noted that the cut score of the test is another factor to be considered concerning the mastery and more attention should be paid to setting up an appropriate cut score.

Discussion

Critique of Task Performance

Overall, it is assumed that the test items will be performed as I have intended. The three writing tasks are appropriate for students because the writing topics such as shopping or cell phones and the functions such as giving advice, expressing their preference, and summarizing a dialogue are closely related to their daily life. In addition, students have learned how to write a paragraph and an email in class for a whole semester, so it is assumed that the test items are proper for summative evaluation. However, considering the students' English proficiency with a wide range from beginner to advanced, I need to elaborate test items for multi-level classes, especially for beginners. I'm afraid the writing items are only three, which are too small to explain and infer the test usefulness.

Evaluation of Test Usefulness

Reliability. The method for reliability is inter-rater reliability, which measures the consistency of ratings by the two raters. It is assumed that reliability will be .75~.85 because the two raters are professionally trained about grading using agreed-on rubrics and they share ideas and negotiate an appropriate score when they have difficult tasks to judge on their own. Meanwhile, internal consistency that is derived from the correlations between items on the test is not measured because the 3 takss are not sufficient to be measured.

Construct related evidence for validity. One of the major considerations of validity is the construct: "How well performance on the assessment can be interpreted as a meaningful measure of some characteristics of quality" (Miller et al., 2008, p.74). Construct validity typically includes content-related evidence and may include criterion-related evidence as well as other types of information (p.103). First, when the TLU domain and the test tasks are compared,

they show the similarities in terms of content-related evidence. It demonstrates that the test tasks have authenticity. In addition, the CBT-type writing test is different from the paper-based writing test in that the CBT-type test includes integrative test tasks, such as reading and writing, and listening and writing; the directions for tasks and score-reporting card written in English. Furthermore, the CBT-type test is composed of three writing tasks for 45 min, while the previous paper-type test has only one writing task for 15 min. Therefore, thanks to authenticity and the characteristic of the CBT-type test, it is assumed that the CBT-type test also has high face validity. Second, I think I can get evidence for construct validity by comparing the students' writing test scores with their previous paper-based writing test scores and previous semester's English grades in terms of criterion-related evidence.

Consequential evidence for validity. Consequences are another consideration in validity through evaluating the effects of use of assessment results on teachers and students. It involves achieving intended positive purposes and avoiding unintended negative effects (Miller et al., 2008, p.74). Though the CBT-type test is more time-consuming and complicated to design and administrate than paper-type tests, it is assumed that the CBT-type test will contribute to increasing the students' motivation, language ability and communicative competence by means of a new format focusing on authentic integrative tests and the grading process might be fair thanks to the well-defined rubrics and two inter-raters.

However, students and teachers are still accustomed to traditional paper-type tests, so it is assumed that some might feel uncomfortable with the CBT-type test. It seems to be consequential negative effects during the transition from the paper-type to the CBT-type test and teachers need to try to avoid unintended negative effects through continuous feedback from students and teachers.

Practicality. It seems that the CBT-type is not a feasible test in terms of practicality. First, teachers are required to acquire knowledge about technology including a computer program for developing a test from creating and editing the tasks, through administrating the test and scoring to issuing a report card. Therefore, it might take more effort and time for teachers to finish all the process of testing from designing the test to issuing a report card. In addition, if some unexpected computer problems that a teacher cannot handle properly occur during a test, it might cause another problem. Second, it needs a computer lab with a networking system for the CBT-type test where at least 40 students can take a test at a time.

Overall estimation of whether or not the test achieved its purpose

Overall, it seems that the test will achieve its purpose in that the test will be used for assigning a grade at the end of a semester; it will provide students with what they can and cannot perform in the given tasks through a report card; it will measure students' integrative writing ability; it will enhance students' motivation, language ability and communicative competence. It is assumed that the test will have evidence for inter-rater reliability and for construct validity as described in Evaluation of test usefulness.

Reflection on personal significance of test

When I look back on developing the test, I have reflected on what I did with the test in the past. To be honest, I used to sometimes develop a test without the TOS if I did not have enough time for the test. Regarding test results and their interpretations, I have paid attention only to the mean and the SD. Unfortunately, less attention has been paid to reliability and validity,; however, now I understand how they contribute to creating much better test tasks.

In addition, I realize that I need to consider the following things to develop a better test. First, when I develop a test, I need to consider the students' perspective more. I learned that I

should not think that students will know what I intend from my directions or prompts. In other words, I need to elaborate the test, paying extra attention to clear directions, comprehensible words in a task or report card. Second, I realized that describing a TLU and making a TOS for assessing is very important. If these were not described precisely and appropriately, the purpose of the test as well as the test results would be meaningless. Last, making rubrics is another work I need to elaborate. I recognized that making well-defined and agreed-on rubrics play a vital role in testing, especially rubrics that are involved with performance tests with a criterion-referenced interpretation. Language testing in practice from designing to issuing a report card includes much more complicated processes than I thought. I try to make a better language test with consideration of the items dealt with in this project.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language assessment in practice*. Oxford: Oxford University Press.

Brown, J.D. (2003). Questions and answers about language testing statistics: Criterion-referenced item analysis (The difference index and B-index). *The JALT testing & evaluation SIG newsletter, 7*(3), pp 18-24.

Connor, U. (2011). *Intercultural rhetoric in the writing classroom*. Ann Arbor: University of Michigan Press.

Higgins, B., Miller, M., & Megmann, S. (2006). Teaching to the test…not! Balancing best practice and testing requirements in writing. *The Reading Teacher*, *60*(4), 310-319.

Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Jeong, H. (2012). *A comparison and analysis of the NEAT and the high school English textbooks-Focusing on writing activities*. Kyunghee University, Seoul.

Jin, Y. (2012). *An analysis of Writing Tests of National English Ability Test: On the basis of communicative functions and languages forms according to the revised 7th curriculum*. Korea University, Seoul.

Joe, M. (2011). *Impact of the KSAT on the EFL environment in Korea*. Woosong University, Jeonju.

Kaplan, R. (1966). Cultural thought patterns in inter-cultural education. *Language Learning, 16*(1&2), 1-20.

Korea Institute for Curriculum and Evaluation. (n.d). Introduction to the NEAT. Retrieved from http://www.kice.re.kr/contents.do?contentsNo=30&menuNo=216

Miller, M. D., Linn, R., & Gronlund, N. (2008). *Measurement and assessment in teaching* (10th

    ed.). Upper Saddle River, N.J.: Prentice Hall.

Stoynoff, S. & Chapelle, C. (2005). *ESOL tests and testing*. Alexandria, VA: TESOL.

Subkoviak, M.J. (1988). A practioner's guide to computation and interpretation of reliability

    indices for mastery tests. *Journal of Educational Measurement, 25*(1), 47-55.

Appendix A
TLU Task Characteristics
TLU tasks: Korean students are taking Language Arts in an English speaking country.

Characteristics of the setting	
physical characteristics	Classroom in an English speaking country
participants	Students, instructor
time of task	Daytime
Characteristics of the test rubric	
instructions	
language	English
channel	Aural and Visual
specification of procedures and tasks	Brief (without examples)
structure	3 Writing Tasks
time allotment	45 minutes
scoring method	
criteria for correctness	Analytic scoring rubric
procedures for scoring the response	1~3 scale, graded by the instructor
explicitness of criteria and procedures	Given, specific
Characteristics of the input	
format	
channel	Aural and Visual
form	Language
language	English
length	3~4 sentences / Paragraph
type	Prompt
degree of speededness	Somewhat speeded
vehicle	Reproduced
language of the input	
language characteristics	
organizational characteristics	
grammatical	Vocabulary(some slang and idioms), syntax (use of various grammatical structure)
textual	Cohesion and rhetorical organization
pragmatic characteristics	
functional	Ideational and instrumental
sociolinguistic	Formal and informal register and naturalness
topical characteristics	Personal
Characteristics of the expected response	
format	
channel	Written
form	Language
language	English
length	80~120 words per each task
type	Extended production response
degree of speededness	Speeded
language of the input	
language characteristics	
organizational characteristics	
grammatical	Vocabulary (some slang and idioms) and syntax (use of various grammatical structure)
textual	Cohesion and rhetorical organization
pragmatic characteristics	
functional	Ideational and instrumental
sociolinguistic	Formal and informal register and naturalness
topical characteristics	Personal
Relationship between input and response	
reactivity	Non-reciprocal
scope of relationship	Broad
directness of the relationship	Direct & Indirect

Test Task Characteristics

Characteristics of the setting	
physical characteristics	Computer lab
participants	Students, instructor
time of task	Daytime
Characteristics of the test rubric	
instructions	
language	English and Korean
channel	Aural and Visual
specification of procedures and tasks	Lengthy (with examples)
structure	3 Writing Tasks
time allotment	45 minutes
scoring method	
criteria for correctness	Analytic scoring rubric
procedures for scoring the response	0~4 scale, graded by two teachers
explicitness of criteria and procedures	Given, specific
Characteristics of the input	
format	
channel	Aural and Visual
form	Language
language	English
length	3~4 sentences / Paragraph
type	Prompt
degree of speededness	Somewhat speeded
vehicle	Reproduced
language of the input	
language characteristics	
organizational characteristics	
grammatical	Vocabulary(some idioms), syntax (use of various grammatical structure), clear articulation
textual	Rhetorical organization
pragmatic characteristics	
functional	Ideational and instrumental
sociolinguistic	Formal and informal register and naturalness
topical characteristics	Personal
Characteristics of the expected response	
format	
channel	Written
form	Language
language	English
length	60~80 words per each task
type	Extended production response
degree of speededness	Speeded
language of the input	
language characteristics	
organizational characteristics	
grammatical	Vocabulary (some idioms) and syntax (use of correct grammatical structure)
textual	Rhetorical organization
pragmatic characteristics	
functional	Ideational and instrumental
sociolinguistic	Appropriate language use
topical characteristics	Personal
Relationship between input and response	
reactivity	Non-reciprocal
scope of relationship	Broad
directness of the relationship	Direct & Indirect

Appendix B

Table of Specifications

	Grammatical Knowledge (vocabulary & language use)	Textual Knowledge (organization)	Functional Knowledge (ideational/ instrumental: giving advice, expressing one's preference, summarizing a dialogue)	# task	# points	% of points
Restrictive Response 1 (giving advice)	0~4	0~4	0~4	1	0~12	33.3
Restrictive Response 2 (Writing one's preference about shopping style)	0~4	0~4	0~4	1	0~12	33.3
Restrictive Response 3 (Summarizing a spoken dialogue about daily life)	0~4	0~4	0~4	1	0~12	33.3
# of task	1	1	1	3		
# of points (the average score of 2 raters)	0~12	0~12	0~12		36	
% of points	33.3	33.3	33.3			100

Appendix C

Copy of the Test (on the computer)

Name	Jeong Sookhee
ID	829597000

Task 1. You will read an email sent by your friend Jason. After reading it, write a reply giving Jason advice about the problem/issue that he writes about in the email..

WHAT'S NEW | INBOX (270) | CONTACTS | CALENDAR | Did you catch the lat... | [Hmart.com] 봄프로 ...

Compose Message | ▼ | Delete | Reply | ▼ | Forward | Spam | 🗎 ▼ | Print | ⚙ ▼ | ↑ | ↓

CBT-TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

Task 1. This is the email sent by your friend Jason. Read it and write a reply giving him a proper advice.

WHAT'S NEW | INBOX (270) | CONTACTS | CALENDAR | Did you catch the lat... | [Hmart.com] 봄프로 ...

Compose Message | ▼ | Delete | Reply | ▼ | Forward | Spam | 🗎 ▼ | Print | ⚙ ▼ | ↑ | ↓

Dear 00,

I have a younger brother who is handsome, smart and popular. When we're together, everybody compares us and he gets all the attention. My aunt even used to say, "How are you so different though you are brothers?", which hurt me. Honestly, I want to be popular like him. Compared to him, I'm average in looks and in school grades. I love my brother, but I don't want to hang out with him because I feel small when I'm with him. Tomorrow we are supposed to go to our aunt's house. What should I do?

Jason

Answer

CBT-TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

Task 1

| WHAT'S NEW | INBOX (270) | CONTACTS | CALENDAR | Did you catch the lat... | [Itmart.com] �ᄅᄅ ... |

Compose Message ▼   Delete Reply ▼ Forward Spam 🗁 ▼ Print ⚙ ▼ ↑ ↓

[Type your response here]

See the text again   NEXT

Time 15:00 / WORD COUNT : 0 /80   CBT-TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

🔊

Task 2. Listen to a question about your shopping style and write your answer with 2 reasons and each reason should have 1 supporting sentence.

Answer

CBT-TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

Task 2

[Type your response here]

NEXT

Time 15:00 / WORD COUNT : 0/80   CBT–TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

Task 3. Listen to a conversation between a man and a woman. Summarize what the woman prefers, sending text messages or making calls, with supporting sentences.

Answer

CBT–TYPE WRITING TEST

Name	Jeong Sookhee
ID	829597000

Task 3

[Type your response here]

FINISH

Time 15:00 / WORD COUNT : 0 /80   CBT-TYPE WRITING TEST

Thank you.

Your responses were sent successfully.

Stay quiet in your place until the teacher lets you

leave.

(수고하셨습니다.
정답이 성공적으로 저장 및 전송되었습니다.
자리에 앉아서 선생님의 지시를 따르세요.)

Writing Test

Appendix D

Sample Answers and Scoring Rubric

1. Sample Answer and Scoring Rubric for Task 1.

(1) Sample Answer for 4 Points

Hi, Jason. I'm so sorry to hear that you feel that way. Though I am your friend, I don't know what you are thinking nowadays. You said that you felt small in front of your brother and you wanted to be popular like him. If you really want to be popular like him, I advise you to smile a lot and have confidence in yourself. I understand how you feel, but don't compare yourself with others. I think everybody is special in his or her own way. Remember, Jason! You are special for just the way you are now. Go to your aunt house with a big smile and confidence in yourself.
Best,

Sookhee

(2) Rubric (adapted from Analytic Essay Rubric for Classroom Use based on *NorthStar* 3 Writing Rubric and ESL composition profile (Jacobs et al. 1981))

	Grammatical Knowledge (vocabulary & language use)	Textual Knowledge (organization)	Functional Knowledge (instrumental: giving advice)
4	-Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Sophisticated range of effective word choice and usage -Some language use and vocabulary errors throughout	-Effective use of transition words to sequence and organize information	-Giving a strong advice with a clear main idea -Relevant to assigned topic with multiple supporting examples
3	-Consistent, generally correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Adequate range of word choice and usage -Several language use and vocabulary errors throughout, but meaning not obscured	-Somewhat effective use of some transition words to sequence and organize information	-Giving a relatively strong advice with a main idea -Mostly relevant to assigned topic with a few supporting examples
2	-Some consistent use of grammar (tense, number word order/function, article, pronouns, preposition) -Limited range of word choice and usage -Numerous language use and vocabulary errors per sentence, and meaning is a little obscured	-Lack of effective use of some transition words to sequence and organize information	-Partially addresses giving advice -Limited supporting examples

1	-Needs Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Very limited range of word choice and usage -Numerous vocabulary and grammatical errors per phrase, and meaning is obscured	-Hard to understand due to no sequencing and organizing information	-Not giving an advice -Not relevant to assigned topic -No supporting examples
0	No response	No response	No response

2. Sample Answer and Scoring Rubric for Task 2.

(1) Sample Answer for 4 points

Even though there are many advantages of buying things online, I prefer to buy things at a store. First of all, I can try and see the actual products if I buy things at a store. To be specific, when buying clothes, it is important to try it on so that I can see if the color matchers me and the size is good. For the second reason, I can get the product right away. It means I don't need to wait until it gets delivered. Lastly, the quality of the product at a store is better than online. I saw many friends complaining about the product purchased online due to bad quality. That's why I like buying things at a store than online.

(2) Rubric (adapted from Analytic Essay Rubric for Classroom Use based on *NorthStar* 3 Writing Rubric and ESL composition profile (Jacobs et al. 1981))

	Grammatical Knowledge (vocabulary & language use)	Textual Knowledge (organization)	Functional Knowledge (ideational / instrumental: expressing one's preference)
4	-Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Sophisticated range of effective word choice and usage -Some language use and vocabulary errors throughout	-Effective use of transition words to sequence and organize information	-Strong preference with two supporting examples -Relevant to assigned topic
3	-Consistent, generally correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Adequate range of word choice and usage -Several language use and vocabulary errors throughout, but meaning not obscured	-Somewhat effective use of some transition words to sequence and organize information	-Somewhat strong preference with two supporting examples -Mostly relevant to assigned topic

2	-Some consistent use of grammar (tense, number word order/function, article, pronouns, preposition) -Limited range of word choice and usage -Numerous language use and vocabulary errors per sentence, and meaning is a little obscured	-Lack of effective use of some transition words to sequence and organize information	-Partial preference with one supporting example. -Limitedly relevant to assigned topic
1	-Needs Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Very limited range of word choice and usage -Numerous vocabulary and grammatical errors per phrase, and meaning is obscured	-Hard to understand due to no sequencing and organizing information	- No preference - Not relevant to assigned topic
0	No response	No response	No response

3. Sample Answer and Scoring Rubric for Task 3.

(1) Sample Answer for 4 points

> *She prefers texting to making calls people. She thinks making calls is time consuming and texting is easier to do and saves time. For example, when she makes calls people to deliver her main message, she needs to spend time on dialing someone's number, saying "Hi" and closing, too. In short, she likes texting because she can text necessary message.*

(2) Rubric (adapted from Analytic Essay Rubric for Classroom Use based on *NorthStar* 3 Writing Rubric and ESL composition profile (Jacobs et al. 1981))

	Grammatical Knowledge (vocabulary & language use)	Textual Knowledge (organization)	Functional Knowledge (ideational / instrumental: summarizing a dialogue)
4	-Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Sophisticated range of effective word choice and usage -Some language use and vocabulary errors throughout	-Effective use of transition words to sequence and organize information	-Summarizing the woman's preference clearly -Relevant to assigned topic with specific supporting examples.

3	-Consistent, generally correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Adequate range of word choice and usage -Several language use and vocabulary errors throughout, but meaning not obscured	-Somewhat effective use of some transition words to sequence and organize information	-Summarizing the woman's preference adequately - Mostly relevant to assigned topic with supporting examples
2	-Some consistent use of grammar (tense, number word order/function, article, pronouns, preposition) -Limited range of word choice and usage -Numerous language use and vocabulary errors per sentence, and meaning is a little obscured	-Lack of effective use of some transition words to sequence and organize information	-Summarizing the woman's preference limitedly. -Limitedly relevant to assigned topic
1	-Needs Consistent, correct use of grammar (tense, number word order/function, article, pronouns, preposition) -Very limited range of word choice and usage -Numerous vocabulary and grammatical errors per phrase, and meaning is obscured	-Hard to understand due to no sequencing and organizing information	-No summarizing the woman's preference -Not relevant to assigned topic
0	No response	No response	No response

Appendix E

Test Material

1. Script for Task 1

> Dear 00,
>
> I have a younger brother who is handsome, smart and popular. When we're together, everybody compares us and he gets all the attention. My aunt even used to say, "How are you so different though you are brothers?", which hurt me. Honestly, I want to be popular like him. Compared to him, I'm average in looks and in school grades. I love my brother, but I don't want to hang out with him because I feel small when I'm with him. Tomorrow we are supposed to go to our aunt's house. What should I do?   Jason

2. Script for Task 2

> **Question**
>
> Some people like shopping on the Internet, others like shopping at a store.
>
> Which do you prefer? To buy things online, or to buy things at a store?
>
> Type your answer with two reasons for your choice and include one supporting sentence for each of your reasons.

3. Script for Task 3

> M: Hi, Lisa. Thank you for texting the message about homework.
>
> F: Not at all.
>
> M: You seem to prefer texting messages than making calls, don't you?
>
> F: Yes, I do. It is easier to do and it saves time. What I mean is that I don't need to spend too much time on dialing someone's number and waiting until he or she answers the phone, saying "Hi", delivering the main message and then closing the phone call which can consume a lot of time. All I have to do is text the necessary message with a short greeting.
>
> M: In my case, making calls is easier to do and it saves time. I can more easily clarify communication gaps and misunderstandings over the phone than by text messages, and it also takes me more effort to send messages than to dial someone's number.
>
> F: You know, it's a matter of individual preference.

Appendix F

Directions for Tasks

The handout about directions will be informed beforehand and students will see and listen to the same directions on the computer in an actual test again.

Ⅰ. General	·Check that the headset functions well. (헤드폰 확인)
	·Enter your name and ID to log in. (로그인 시 이름, ID 입력)
	·Click "Answer" when you write an answer. (답안 작성시 answer 클릭)
	·When you finish each task, click "NEXT" (When the assigned time passed, the task will go "NEXT" automatically and you cannot go back to "Previous" task. (다음 문항 이동시 "NEXT" 클릭, 이전 문항으로 이동 불가)
	·When you finish the last task, stay in your seat quietly. (시험 종료 후 조용히 자리에 앉아있기)
	·Your answer will be saved and sent to the teacher automatically. (답안은 자동 저장 및 자동 제출)
	·Time left and word counts are shown on the monitor. (종료시간, 단어수는 모니터에 표시됨)
	(EX) 11: 30 means you have 11 min. and 30 sec. left to complete your answer. (11:30 은 시험 종료시까지 11분 30초 남았음을 의미함)
	(EX) Word counts: 75 means your answer is composed of 75 words. (word counts :75 는 현재 75 단어를 사용했음을 의미함)

Ⅱ. Content of the Test	·The test is made up of three writing tasks as follows;
	(시험은 3가지 유형으로 구성되어있음)
	(1) reading a text and writing a reply giving advice (60~80 words / 15 minutes)
	(주어진 글 읽고 충고하는 글 쓰기, 60~ 80단어/15분)
	(2) writing your preference after listening to a question (60~80 words / 15 minutes)
	(선택의문문을 듣고 자신이 선호하는 것을 쓰기, 60~80 단어/15분)
	You can listen only once. (문제 내용은 한 번만 들을 수 있음)
	(3) summarizing a spoken dialogue about daily life (60~80 words / 15 minutes)
	(대화문 듣고 요약하기, 60~80 단어/15분)
	You can listen only once. (문제 내용은 한 번만 들을 수 있음)
	☞ You can use the given scratch paper for taking notes.
	(메모를 위해 주어진 종이를 사용할 수 있음)
	☞ The writing should consist of 60~80 words, or your score will be deducted.
	(60~80 단어를 사용해야 함. 위반 시 감점됨)
Ⅲ. Scoring	· Your task will be evaluated by two raters in terms of grammatical knowledge (vocabulary & language use) , textual knowledge (organization), and functional knowledge (ideational / instrumental).
	(채점은 공동 채점하며, 채점 영역은 문법지식, 텍스트지식, 기능지식)

Appendix G

Score-reporting Form

The individual writing score will be uploaded on school homepage and students can check their score on the computer and print it out anytime.

Writing Score Report Card

Name	
ID	

Task	Test Date Nov. 7, 2012	Points earned	Your performance was... (Eng./Kor.)	Sub total
Task 1	Grammatical knowledge	__ points		
	Textual knowledge	__ points		___ / 12
	Functional knowledge	__ points		
Task 2	Grammatical knowledge	__ points		
	Textual knowledge	__ points		___ / 12
	Functional knowledge	__ points		
Task 3	Grammatical knowledge	__ points		
	Textual knowledge	__ points		___ / 12
	Functional knowledge	__ points		
Overall Comments (Eng./Kor.)				Total Score __ / 36

☞ The cut score is 18 / 36 (50%).

CSU HIGH SCHOOL