

Sociopolitical and Contextual Influences on Teacher-produced Achievement Tests of English in Korean High Schools

Min-Chang Sung*
Seoul National University

Victoria H. Jo
Seoul National University

Abstract

The increasing importance of classroom-based assessment has resulted in extensive studies of teacher-produced tests in a variety of foreign language learning contexts. While this area of research has established a substantive procedural framework for constructing teacher-produced tests, not much effort has been directed towards exploring the quality as well as the relevance of actual testing materials that are developed by local schoolteachers. One of the key issues to consider in analyzing teacher-produced tests is whether the test is designed to measure what it purports to measure. Employing both quantitative and qualitative approaches, the present study first explored the distinctive characteristics of teacher-produced English language achievement tests at eight Korean public secondary schools and conducted semi-structured interviews with teachers to determine the factors influencing the constructs of the tests. Results indicate that the constructs of teacher-produced English language achievement tests are profoundly affected by dominant external forces—namely, a high priority given to college admissions test preparation as well as a range of limitations imposed by policy regulations and restrictions.

* Corresponding author (potamin3@snu.ac.kr)

Keywords: English language testing, classroom-based assessment, teacher-produced achievement test, Korean public high school, washback effect

I. Introduction

Classroom-based assessment (hereafter, CBA) has been extensively investigated in the field of foreign language education (Black & William, 1998; Brindley, 1998, 2001; Clarke & Gipss, 2000; Coniam, 2009; Davison, 2004; Hill, 2012; Hill & McNamara, 2011; Jeon et al., 2005; Jeong, 2012; Leung, 2005, 2007; Nitko, 1989; Rea-Dickins, 2001, 2006, 2007). The important role of the CBA in foreign language teaching and learning has been discussed from a variety of perspectives: the outcome of the CBA helps the teacher to make pedagogical decisions (Nitko, 1989); the CBA has a direct impact on the way learners study foreign languages (Black & William, 1998; Leung, 2007; Rea-Dickins, 2001, 2006); high-quality CBA encourages learners to gain positive attitudes towards instruction (Madsen, 1983); and the scores in the CBA are used as a principal indicator of the learners' foreign language proficiency (Coniam, 2009; Jeong, 2012).

Given the increasing importance of the CBA in foreign language learning contexts, a number of researchers (e.g., Coniam, 2009; Hill & McNamara, 2011; Jeong, 2012; Rea-Dickins, 2001) have placed a special focus on a major type of the CBA: the teacher-produced achievement test (hereafter, TPAT). The analysis of testing materials developed by foreign language teachers has revealed the complex nature of the TPAT. Hill and McNamara (2011), for example, examined assessment processes in two Indonesian language classes and argued that it is necessary to explore the diverse nature of the TPAT from an emic perspective (i.e., insiders' perspective). Similarly, Rea-Dickins (2001) explored the complex and multifaceted aspects of the TPAT and contended that, "when we try to answer the question '*What is assessment?*'", there is no one right or easy answer" (p. 458).

The complex nature of the TPAT in foreign language classes is particularly evident at high schools in South Korea, as a variety of

sociopolitical issues are closely related to English language teaching and testing. First, the English language is taught as one of the core subjects alongside with mathematics, literature, and science; therefore, TPATs of English are administered as part of the midterm and the final exam among different sets of TPATs in ten to twelve subjects, each lasting approximately an hour over the course of three to five days. Second, the assessment of English language proficiency heavily relies on the TPAT, and thus the students' scores in the TPAT are regarded as one of the two most important criteria which count toward university admission qualification, the other being the College Scholastic Aptitude Test (CSAT) scores. Finally, in recognition of the vital role of TPAT in English language teaching and learning, the Korean Ministry of Education has developed a series of policy guidelines for the national curriculum of English language subject, commissioning the local teachers of English to produce high-quality TPATs.

Despite these various contextual influences, however, most of previous studies that diagnosed significant problems in TPATs of English at Korean secondary schools placed little attention to identifying types of influential sociopolitical factors (Byun, 2002; Na, 2012). For example, Gwack (2008) analyzed TPATs of English at thirty high schools based on the framework for test method facet (Bachman and Palmer, 1996) and identified several problems in the tests. The author further suggested six types of solution, but most suggestions were in their nature remedial and restricted to reducing errors found in the tests. Other studies (Im, 1995, 1996) investigated local English teachers' testing practices via a questionnaire survey and noted the lack of teacher professionalism as a primary cause of problems in TPATs of English. These studies seem to have been guided by such questions as "what are major problems in TPATs?" or "how can we solve the problems?", without addressing fundamental adverse factors that local schoolteachers encounter when developing and implementing TPATs of English.

Accordingly, the present study aims to uncover the major source of the problem by investigating the distinctive characteristics of recent TPATs of English. Employing both quantitative and qualitative approaches, this study examines the determining sociopolitical and

contextual factors in the development and administration of the TPATs. The main research questions of the present study are as follows:

1. What are the distinctive characteristics of the TPATs of English?
2. What are the most dominant sociopolitical and contextual factors which adversely affect the TPATs?

In line with these questions, the present study further discusses whether certain conditions are required for the effective management of the factors which adversely affect the TPATs

II. Literature Review

A. Definition and characteristics of TPAT

According to Hill and McNamara (2011), CBA is broadly defined as “any reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes” (p. 396). The broad definition of CBA includes different types of tests (Bachman, 1990) which vary in terms of *purposes* (e.g., diagnostic and achievement); *contents* (e.g., general proficiency and formative); *scoring* (e.g., subjective and objective); and *methods* (e.g., performance and multiple-choice).

Among the varieties of CBA, the critical role of TPATs in EFL contexts has been extensively explored in the literature (Coniam, 2009, 2014; Gronlund, 1985; Jeong, 2012; Nitko, 1989). For instance, Gronlund (1985) suggests that, in comparison to standardized tests, TPATs may be more relevant as a measure of the test-takers’ achievement levels—even though the standardized test can generally claim better validity and reliability than the TPAT. In a similar sense, Nitko (1989) noted that the primary purpose of the TPAT is two-fold: to provide teachers and students with feedback on the effectiveness of a particular foreign language course, and to help them make informed

pedagogical decisions for future teaching and learning. Typically, TPATs are administered at the end of instructional phase in order to include all contents taught in class leading up to the test period (Bachman, 1990; Black & William, 1998).

B. English language assessment in South Korea

Teaching the English language as a subject matter in Korean secondary schools commenced in 1945 (Jung & Norton, 2002). From the time of its inception as a regular subject, English language classes have used individual student scores on the TPATs as the important indicator of academic achievement, despite the mounting criticism of its validity, such as mismatches between course objectives and testing items, a very small number of items testing learners' integrated skills, and students' negative attitudes toward the test (e.g., Im, 1995, 1996; Jeong, 2012). In particular, the discrete-point testing methods (e.g., multiple-choice, matching, true-false, and fill-in-the-blank), which are most widely used in TPATs, are held to be invalid measures for properly assessing learners' English language skills (Jung 2007; S. Kim, 2009). Even the learners seem to remain skeptical about the relevance of the discrete-point test items to successful English language learning (Choi, 2008).

Within the English language education domain in South Korea, the aforementioned problems inherent in using TPATs have been often attributed to the teachers' limited knowledge of proper assessment techniques, combined with lack of adequate practical experience in test development (Im, 1995, 1996; Jeong, 2012). However, other researchers who have investigated the assessment practices in diverse foreign language contexts take a very different view on the teacher's role in the development of TPATs—that is, the teacher is recognized as a specialist, qualified to develop the most appropriate testing module for students (Hill & McNamara, 2011; Y. Kim, 2009; Rea-Dickins, 2001; Shih, 2010). The focus of this line of research was on the analysis of dynamic processes involved in planning, administration, and interpretation of assessment. For example, Hill and McNamara (2011) investigated two Indonesian language classrooms in Australia in the

process of implementing a new curriculum. Exploring a variety of data (i.e., audio-recordings, classroom observation, field notes, and school documents), the authors provide a detailed description of dynamic processes involved in the CBA. Similarly, Shih (2010) explored the complex nature of CBA in two Taiwanese universities, specifically focusing on how these institutions are in the process of determining a new policy direction for the implementation of a standardized English test as a requirement; identifying several influential factors such as educational, school, and parental factors, the author developed a model of washback on policies at schools. In addition, Rea-Dickins (2001) and Y. Kim (2009) noted that contextual variations such as teacher's L1 background, lesson focus, and class size synthetically impact assessment practices.

With the conceptual framework mentioned above, the aim of the present study is to present descriptive data on the main distinctive characteristics of the TPATs that have been developed for English classes at eight high schools in South Korea. Moreover, this study is conducted to determine whether there are dominant sociopolitical and/or contextual factors that may adversely affect the development and administration of the TPATs, and if so, what is required to effectively manage the complexity of external influences that shape the outcome. The practical effectiveness and relevance of TPATs to successful English language learning are discussed from a pedagogical perspective.

III. Methods

The present study conducted two types of in-depth analysis. First, the test items in recent TPATs of English from eight Korean high schools were analyzed using the following four criteria: a) test-item formats, b) skills or knowledge tested, c) types of stimuli, and d) test construction errors. These four criteria have been frequently used when describing distinctive characteristics of TPATs for English (Im, 1995, 1996; Kwak, 2008; Na, 2011). Second, a questionnaire survey and semi-structured individual interviews were carried out with the teachers who developed the tests.

A. Data collection

Eight public high schools in Seoul and its vicinity were randomly chosen. None of the schools are special types of high school such as foreign language high schools, science high schools, or autonomous private high schools; therefore, it can be said that all of the schools analyzed in the present study are categorized as *normal* high schools. These schools seem to represent that type of high schools for the following two reasons: a) the pool includes male, female, and mixed schools, and b) the pool covers representative educational sections such as Gangnam, Gangbook, and Gyeonggi areas.

The schools are evenly balanced into two groups according to target grade and time of test administration. Table 1 provides the specifications of the tests collected from the two groups.

Table 1. Tests collected from eight high schools grouped by grade levels

School	Grade	Year	Semester	Test
A, B, C, D	10th	2012	Spring	Midterm exam
E, F, G, H	11th	2013	Fall	Midterm exam

The collected tests from the first group were TPATs of 10th-grade English which were administered as midterm exams in the spring semester of 2012; the tests collected from the second group were TPATs of 11th-grade English which were administered as midterm exams in the fall semester of 2013. The present study excluded the analysis of final exams, as midterm and final exams have the same purpose, which is to test what the student has learned since the end of the previous achievement test. The only difference between midterm and final exams is the time of administration.

The TPATs of 12th grade English are also excluded from the present study because the tests that are given to high school seniors (i.e., 12th graders) are largely based on the CSAT study materials developed by a public publishing company, from which English reading texts are extracted to construct as much as 70 percent of items in the actual CSAT

that will be taken by the seniors (Korea Institute for Curriculum and Evaluation, 2012).

B. Data analysis: Quantitative study

All test items in the eight TPATs are categorized into the following four components.

Test-item formats (multiple-choice, controlled-production, & free-production)

According to Brown (1996), test-item formats must be appropriate for the skills or knowledge areas tested. For instance, using multiple-choice items to measure productive skills (i.e., speaking and writing) is ineffective. In the present study, test items are categorized as one of the following types: multiple-choice, controlled-production, and free-production (Allen & Spada, 1983). Multiple-choice items ask test-takers to select one or more answers, whereas the other two test-item types (controlled- and free-production) ask test-takers to write or utter the answers. Specifically, the controlled-production item examines whether the test-taker is able to produce the correct answer, whereas the free-production item focuses on the overall comprehension of a given material as well as the ability to formulate a logical, coherent, and creative answer (Allen & Spada, 1983; Bachman, 1990).

Skills (reading, writing, listening & speaking) and knowledge (grammar & vocabulary) tested

Distinct from commonly used standardized language tests (e.g., TOEFL and IELTS) in which each section is designed to assess a single skill area (i.e., speaking or writing), the TPATs are geared to measure a composite of multiple discrete components, covering varied skill and knowledge areas pertaining to the subject matter. For the purpose of the present study, this necessitated a further classification of the skill and knowledge areas: each test item was classified into one of the four language skills (reading, writing, listening, and speaking) and/or two knowledge areas (grammar and vocabulary). In a few cases where the skill and knowledge areas overlapped, the test item was categorized into

the skill/knowledge area deemed more essential for answering the question. For example, a test item which consisted of a 150-word text with two fill-in-the-blanks was categorized as a measure of reading skill, rather than that of lexical knowledge, since the test-taker must comprehend the overall message of the text in order to determine the correct words that fit in the blanks. (See the Appendix for sample test items.)

Types of stimuli (non-linguistic, discrete sentences, texts from textbooks or from mock CSATs)

Bachman and Palmer (1996) provided a variety of ways for classifying stimuli types, among which two major distinctions are identified as most relevant for the present study: *form* (linguistic vs. non-linguistic) and *length* (sentence-level vs. discourse-level). The initial search employing these two criteria revealed that the textual stimuli presented in TPATs derived from not only textbooks but mock CSATs (which are administered to high school students every two to three months), even though the contents of such tests are not explicitly taught as part of English language classroom instruction. This particular finding called for a further specification of textual input based on the resource, and thus all of the stimuli in the TPATs were categorized into the following five types: a) non-linguistic input, e.g., drawings and photos, b) discrete sentences, c) texts from textbook, d) texts from mock CSATs, and e) miscellaneous.

Errors (based on evaluation criteria)

Published evaluation criteria on test items (Brown, 1996; Harris, 1969; Im, 1995, 1996; Lado, 1970; Rivers, 1981) were synthesized into the following eight evaluation criteria:

- 1) The item should be in a format appropriate for the testing purpose and content;
- 2) The item should not use double negation;
- 3) One item should not give a hint for other items;
- 4) All relevant information of an item should be presented on the same page;

- 5) Irrelevant information should be excluded;
- 6) Every distracter should sound attractive;
- 7) Unnecessary repetition should be avoided; and
- 8) Miscellaneous: The item should be free of any other problem.

Every item in the TPATs was examined with reference to the above eight evaluation criteria, and only the items that both researchers judged as in violation of the above criteria were identified. Of the set of eight criteria, an additional explanation is provided below for those less self-evident criteria such as the criterion 1 and criterion 5.

Violation of criterion 1: The test item should be in a format appropriate for the testing purpose and content.

※ *Read the following text and correct an ungrammatical expression.*

Oliver is a professional soccer player and the founder of Soccer for Hope, an organization that helps children with severe illnesses. Oliver liked Nicolette's idea and agreed to help set up a bank account for *Futbol 4 Refugees* so that people could easily send donations. Oliver also donated the first 25 soccer balls. Since then, Nicolette promoted *Futbol 4 Refugees* at gatherings of teams that she belonged to as a player and at parents' meetings. She has even contacted a number of well-known companies.

Ungrammatical expression Correct expression

_____ → _____

(Extracted from the test of School B)

Intending to measure grammatical knowledge, the above test item asks students to detect and correct an ungrammatical expression (i.e., *promoted* → *has promoted*) in an 84-word text. As it may be confusing to even native English speakers, this is identified as a test-item in violation of the Criterion 1. An additional issue related to the test item connected to criterion 1 is that the focal sentence is actually 'possible' idiomatically speaking. Arguably, a more appropriate format for testing grammatical knowledge would be a test item with several underlined

phrases from which a test-taker is asked to detect and correct the grammatical error.

Violation of criterion 5: Irrelevant information should be excluded.

※ Choose the word that fits in both blanks.

- To _____ means to carry out, or bring to realization, as a prophecy or promise.
- Introduced as the 17th coach in Flyers history last night, Peter Laviolette said he is here for one reason: to bring the city a hockey championship. “I understand in Philadelphia there’s an expectation for success,” he said during a news conference at the Wachovia Center. “And I’m going to do my best to _____ that expectation and bring a Stanley Cup back to Philadelphia.

- ① approach ② declare ③ fulfill
④ ignore ⑤ represent

(Extracted from the test of School A)

The above test item asks students to find a word that fits in both blanks in the paragraph. As the second stem is unnecessarily long, this is identified as a test-item which is in violation of Criterion 5—*irrelevant information should be excluded*. A more appropriate format would be a simpler sentence without irrelevant information, as in the following example:

The coach said, “I’m going to do my best to _____ the citizens’ expectation and bring a Stanley Cup back to Philadelphia.”

This shortened sentence drastically reduces the number of words in the second stem, from 66 to 22, without any loss of relevant information.

C. Qualitative study: survey & interview

Striking findings from the quantitative analysis of test items were investigated using two qualitative methods. First, a questionnaire of 17 items was devised to examine the local schools’ policies and the

teachers' beliefs and practices concerning TPATs. The questionnaire was provided in the teachers' L1 (i.e., Korean) so that they could freely express their opinions. The teachers' responses in the survey were informative in analyzing the quantitative results; however, for a comprehensive and detailed understanding of the teachers' beliefs and practices, the questionnaire survey was supplemented by semi-structured, one-on-one interviews. The interviewees were the teachers from four schools (C, D, E and F), as the analysis of the TPATs in these schools showed intriguing patterns and/or the teacher's responses in the questionnaire required more in-depth accounts. Therefore, the interview questions were based on the analysis of test items and the teachers' responses to the survey questions. The interviews were recorded under the interviewees' permission.

IV. Results

A. Test-item formats

There are three types of test-item formats: multiple-choice, controlled-production, and free-production item. A multiple-choice item gives several options and asks students to select one or more correct answers, while a controlled-production or a free-production item asks students to produce (i.e., write and/or speak) an answer. The distinction between a controlled-production and a free-production item is based on whether a correct response is to be produced (i.e., controlled-production) or students' overall comprehension and/or language proficiency is measured (i.e., free-production).

The average number of items per test was 33.25 ($SD=5.06$). The TPAT of School G had the greatest number of items (41), whereas the TPAT of School C had the lowest number of items (25). Table 2 shows the number of test items in each format.

Table 2. Item formats in the tests of eight high schools

School	Multiple-choice	Controlled-production	Free-production	Total (Rank)
A	31	8	0	39 (2)
B	28	6	0	34 (3)
C	19	6	0	25 (8)
D	24	6	0	30 (7)
E	28	6	0	34 (3)
F	27	5	0	32 (5)
G	35	6	0	41 (1)
H	25	6	0	31 (6)
Total (%)	217 (81.6%)	49 (18.4%)	0 (0%)	266 (100%)
Mean (SD)	27.13 (4.76)	6.13 (0.83)	0 (0)	33.25 (5.06)

The most frequent format used in the TPATs was the multiple-choice type, accounting for over 80% of the total items, while no free-production items were found in the tests. Notably, the number of controlled-production items was relatively small and consistent among the schools (Mean=6.13; SD=0.83), whereas that of multiple-choice items was relatively great and inconsistent (Mean=27.13; SD=4.76). The greatest gap in the number of controlled-production items was only three (between School A and F); however, the gap in the number of multiple-choice items was about five times higher, i.e., sixteen (between School G and C).

B. Skills and knowledge areas

The category of *skills and knowledge areas* has six components: four language skills (i.e., reading, writing, listening, and speaking) and two knowledge areas (i.e., grammar and vocabulary). The proportion of each component was calculated to identify which skill(s) or knowledge area(s) the TPAT appears to measure. In total, the greatest proportions were observed in reading (34.6%) and vocabulary (23.3%), which accounted for more than a half of the total items. The rest of the test

items appeared to measure students' grammatical knowledge (19.9%), writing (19.5%), listening (1.9%) and speaking (0.8%) skills.

Table 3 presents the numbers of items for each of the skills and knowledge areas in the eight TPATs, along with the means and standard deviations. Although the items on reading were the most frequently observed (Mean=11.5), the greatest standard deviation was in the items on vocabulary (SD=5.5). Most notably, the tests at Schools A and E had 16 items on vocabulary, whereas the test of Schools C only had two items on vocabulary.

Table 3. Distribution of items for skills and knowledge areas by school

School	Read.	Writ.	Listen.	Speak.	Gram.	Vocab.	Total
A	10	6	0	0	7	16	39
B	17	6	0	0	8	3	34
C	9	9	0	1	4	2	25
D	11	3	0	1	9	6	30
E	10	6	0	0	2	16	34
F	11	6	0	0	6	9	32
G	11	6	5	0	13	6	41
H	13	10	0	0	4	4	31
Total	92	52	5	2	53	62	266
(%)	(34.6)	(19.5)	(1.9)	(0.8)	(19.9)	(23.3)	(100)
Mean	11.5	6.5	0.6	0.3	6.6	7.8	33.25
(SD)	(2.5)	(2.1)	(1.8)	(0.5)	(3.5)	(5.5)	(5.06)

C. Formats and skills/knowledge areas

Figure 1 presents the proportions of the two item formats (i.e., multiple choice and controlled-production) in four skill/knowledge areas¹. Items on reading and vocabulary predominantly relied on

¹The free-production item format was excluded because none of the tests used this format. Listening and speaking skills were also excluded from the analysis because the total number of items on these skills in the eight sets of achievement test were only five and two, respectively, thereby, rendering it inappropriate to examine the proportions of each item format.

multiple-choice format, 96.7% and 90.3% respectively, while those on writing employed both multiple-choice format (46.2%) and controlled-production items (53.8%). In addition, every school had at least two controlled-production items on writing.

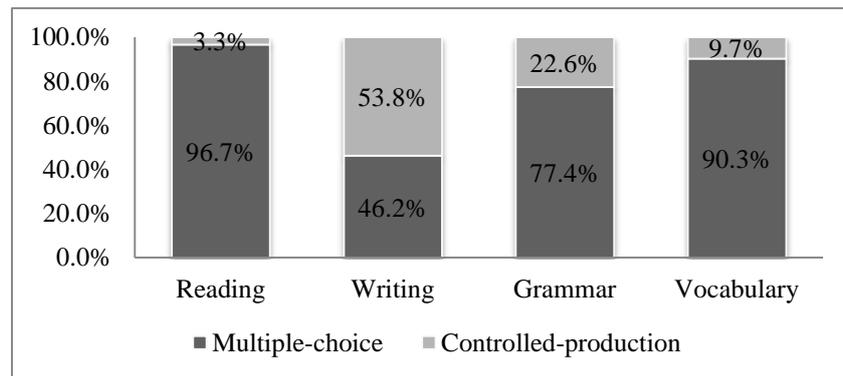


Figure 1. Proportions of item formats for skills and knowledge areas

D. Stimuli

Stimuli refer to (non-)linguistic materials to which test takers are requested to respond. The stimulus of every test item was categorized into the following five types according to its form, length, and source: non-linguistic materials such as drawings and photos, discrete sentence tokens, texts extracted from textbook, texts extracted from mock CSAT, and miscellaneous.

Table 4 shows the proportions of stimulus types for test items, which reveals a great reliance on texts from textbooks or mock CSAT. In total, stimuli of text type accounted for approximately 87% of the test items. Interestingly, non-linguistic materials such as drawings and photos were not provided as stimuli in any of the tests.

Table 4. Proportions of stimulus types in the tests of eight high schools

School	Non-linguistic	English Sentence	Text: Textbook	Text: Mock CSAT	Miscellaneous	Total
A	0.0%	23.1%	38.5%	38.5%	0.0%	100%
B	0.0%	2.9%	79.4%	17.6%	0.0%	100%
C	0.0%	0.0%	84.0%	16.0%	0.0%	100%
D	0.0%	0.0%	96.7%	0.0%	3.3%	100%
E	0.0%	23.5%	29.4%	38.2%	8.8%	100%
F	0.0%	12.5%	71.9%	12.5%	3.1%	100%
G	0.0%	9.8%	22.0%	68.3%	0.0%	100%
H	0.0%	3.2%	25.8%	58.1%	12.9%	100%
Total	0.0%	11.1%	57.0%	29.8%	2.1%	100%

An additional interesting finding is that the tests of School E, G, and H had extracted texts more frequently from mock CSATs than from textbooks. These schools, along with School F, belong to Group 2, from which the 11th-grade tests were collected, while the other four schools (A, B, C, and D) belong to Group 1, from which the 10th-grade tests were collected. As shown in Table 5, there are notable differences in the proportion of stimulus types between the 10th-grade and the 11th-grade tests. Specifically, a significant difference is found in the use of textual stimuli: the primary source of textual stimuli for the 10th-grade tests was textbooks (71.9%), whereas that for the 11th-grade tests was mock CSATs (45.7%).

Table 5. Proportion of stimulus types in the 10th-grade and 11th-grade tests

	Non-linguistic	English Sentence	Text: Textbook	Text: Mock CSAT	Miscellaneous
10th Grade (A/B/C/D)	0.0%	7.8%	71.9%	19.5%	0.8%
11th Grade (E/F/G/H)	0.0%	12.3%	36.2%	45.7%	5.8%

E. Violations of evaluation criteria

The present study has devised eight evaluation criteria for test items on the basis of previous studies in foreign language testing. Every item in the TPATs was examined with reference to the evaluation criteria, and test-items in violation of any of the criteria were identified. Table 6 illustrates the number and distribution of test-items containing violations in the TPATs of English in the eight high schools. The largest number of violations was observed in the test of School A (14 violations), while the smallest number of violations was found in the tests of School C and H (3 violations). The remaining schools' tests contained five to seven violations.

Table 6. Violations of evaluation criteria for test items in eight high schools

Evaluation Criteria	A	B	C	D	E	F	G	H	Sum
1. The item should be in appropriate format.	0	3	1	0	7	1	2	0	14
2. The item should not use double negation.	0	0	0	0	0	0	0	0	0
3. One item should not give hint for other items.	1	0	0	0	0	0	0	0	1
4. An item should be presented on the same page.	3	0	0	3	0	5	3	0	14
5. Irrelevant information should be excluded.	2	1	0	0	0	0	0	0	3
6. Every distracter should sound attractive.	2	3	0	0	0	1	0	0	6
7. Unnecessary repetition should be avoided.	0	0	0	0	0	0	0	0	0
8. The item should be free of any other problem.	6	0	2	2	0	0	1	3	11
Total	14	7	3	5	7	7	6	3	49

Apart from the total number of violations, interesting patterns were observed in different school contexts. First, a vast majority of violations detected in the tests of School D, F, and G seemed to be related to the construct of the test (e.g., presenting an item within a page). Second, every violation observed in the test of School E was due to inappropriate format for the testing purpose and content. For example,

there was a five-option multiple-choice item on vocabulary, where each of the option had five pairs of English words and their Korean translations, showing a total of 25 pairs of words. Students were then asked to choose one option showing an incorrect Korean translation out of the five options, which required them to detect one incorrect pair out of the given 25 pairs of words.

V. Discussion

In an effort to unveil the complex nature of the TPATs in eight Korean high schools, this section discusses major findings from the preceding analysis with reference to the teachers' responses in the questionnaire survey and the interviews. According to the teachers' comments, the constructs of TPATs are significantly affected by both internal and external forces—namely, a range of administrative limitations to testing and the high priority placed on preparing students for the university admissions.

A. Impact of internal factors on TPATs

Limitations imposed by policy regulations and restrictions

The analysis of the interview data revealed that the overall structure of TPAT is largely dictated by the venue and test room facilities, which present major physical limitations to the administration of tests. One important finding in the present study is the absence of free-production tests—that is, none of the eight schools investigated in this study included any free-production test items as an authentic measure of English language proficiency. A primary reason for a complete lack of free-production items in TPATs is that, the schools are required to administer the tests in compliance with policy regulations, according to which all students must take the test at the same time in the same place. Under this restriction, the current lack of access to sufficient recording devices makes it exceedingly difficult, if not impossible, for schools to include free-production tests in the TPATs, which must be held concurrently and at the same venue—under the same conditions, rules,

and time controls. A teacher's view on the non-practicability of including free-production items in the TPAT is noted in the following excerpt:

Teacher D: *It is very difficult to administer a speaking test to the entire student body simultaneously (due to lack of equipment). This is why speaking skills are not evaluated by using the midterm and final test but by other additional tests.*

The finding supports the claim made in Shih (2010), which argued that school factors have a great impact on CBA practices. A notable difference, however, was observed: in Shih (2010), the teachers seemed to play a more independent role, evaluating the suggestions provided by a higher educational institute and deciding whether to follow them, while, in the present study, the teachers seemed less independent, expressing that they have to follow the instruction from a higher institute. For example, when a teacher from School C was asked why she had used controlled-production items to test students' writing skills, she explained that, even in the official guidelines developed by the Metropolitan Office of Education, the sample test items on writing have been constructed in a way that students simply write one or two sentences instead of producing whole paragraphs or essays (Figure 1):

Teacher C: *When you look at a collection of sample test items produced by the Metropolitan Office of Education, you will find that most of the items for testing writing skills are kind of controlled writing such as unscrambling sentences or fill-in-the-blanks.*

Another administrative regulation on production items is that at least 30% of the total score must be allotted to that type of items. This implies that local schools may assign as much as two-thirds of the TPATs to production items; however, this rarely occurs in actuality as local schools are only required to comply with the minimum ratio level.

According to the interview data, the teachers are reluctant to incorporate many production items in their TPATs because such types of test items would impose additional administrative burdens such as scoring individual answers and dealing with students' complaints and grade appeals. This may be the primary reason for the absence of free-production items as well as a small and consistent number of controlled-production items in the TPATs that had been administered across the eight high schools.

Inherent quality problems in TPATs

The overall structure of the tests—the number of items and distribution of item format—varied noticeably among the eight schools investigated. For example, the achievement tests are given three times during a semester at School C, whereas the remaining seven schools customarily administer only two tests each semester. This difference in the frequency of testing during a semester explains School C's relatively smaller number of items per test (25) compared to those of other schools (at least 30 test items). One of the teachers from School C expressed that, although administering three tests each semester may be effective in compelling students to study throughout the semester, the extremely labor-intensive process of generating as well as grading the tests each time places undue pressure on the teachers:

Teacher C: *It is quite burdensome for teachers to develop, administer, and grade tests; there are too many contextual factors to consider, and it takes too much time!*

In addition, the effectiveness of the TPAT in achieving major pedagogical objectives of English courses is questionable. The functionality of the widely-known concept of 'testing for learning' is two-fold: a) the assessment itself induces learners to study what have been taught, and b) the assessment results provide diagnostic information about student progress toward mastery of learning goals which can guide teachers in decisions about how to advance learning. In the present study, the TPATs were being used to serve the former

function exclusively, which adversely affected the overall quality of TPATs.

Another notable factor contributing to quality problems inherent in TPATs is the custom of having teachers take turns reviewing/editing the entire set of tests. In an effort to ease their test-related workload, it is a common practice in local schools among teachers to rotate reviewing/editing responsibility. As reflected in the following excerpt, individual teachers assume the role of editor of the test only when it is their turn to serve:

Teacher F: After developing test items for the portion of the test that have been assigned to us, we submit them to a teacher whose turn it is to serve as editor. We rotate this role among us each time we give a test, and the editor is responsible for reviewing the test prior to administering it to students.

Allocated on the basis of the total number of English teachers at the eight schools (Mean=11.625; Range=9 to 16), a typical teacher may be required to serve as the editor only once in three years. Although such a way of managing TPATs may help to reduce teacher workload, it seldom affords teachers the opportunity to refine their skills in test-construct and editing, which are indispensable skills for schoolteachers of all grade levels and subject areas. These school practices may account for the many problems related to test-construct and editing in the tests of School D, F, and G. Within the hierarchical organizations of teachers, it is also conceivable in some cases that senior teachers may simply delegate their duty to junior members of the teaching staff who are less experienced in test construct. Most seasoned teachers presumably take their editing responsibility with due seriousness; nevertheless, without redesigning the current system of developing TPATs, teachers' lack of expertise in test making and management will be an inevitable outcome.

B. Impact of external factors on TPATs

Preparation for the CSAT

The primary purpose of the achievement test is to measure student progress (Bachman, 1990). Aside from assessing individual student progress, the structure of the achievement tests administered to Korean high schools clearly shows that there is another important function of the TPATs—namely, preparation for the high-stakes CSAT required for entry into higher education. As the CSAT is one of the most important criteria required for admission into universities in Korea, the first and the foremost goal of high school teachers is to prepare students for the CSAT—which is reflected in the common practice of adopting the CSAT format in the TPATs. Most notably, the TPAT of School E contained many items from a vocabulary workbook, despite the fact that the vocabulary items being tested have never been explicitly taught in class. As reflected in the following excerpt, the purpose of including miscellaneous vocabulary items (which was not taught) in the TPAT is to compel students to learn the words that may appear on the CSAT:

Teacher E: *We decided to develop ten items from a vocabulary workbook even though we don't actually teach any vocabulary in this workbook in class. By including some vocabulary from the workbook in the test, we can easily encourage students to learn new words. We believe this is an effective way to prepare them for the CSAT.*

An additional aspect reflecting the washback effect of the CSAT concerns the common practice of using multiple-choice questions based on 150- to 200-word reading texts in the TPAT, which is a prevailing test-item format in the CSAT. Specifically, this washback effect of the CSAT was extremely strong that this type of textual stimuli accounted for 87% of the total test items in the TPATs. Moreover, in an effort to properly prepare students for the CSAT, the teachers at the eight schools regularly adopted reading texts from mock CSATs as a stimulus in their own TPATs. This was evidenced by such comments as: “I develop test

items using stimuli from other important resources besides the textbooks in order to prepare my students for the CSAT” (Teacher F: Translated by the researcher). In particular, the significantly higher proportion of texts extracted from mock CSAT in the eleventh-grade TPATs (45.7%) compared to the tenth-grade TPATs (19.5%) seems to manifest the greater importance of university admissions test preparation for the eleventh-graders who will be taking the CSAT in the following year. This finding is in line with the seminal work by Shohamy et al. (1996), showing that the national test of English has a great impact on teaching and evaluating practices in English classes at local schools.

Mirroring the most prevalent test-item formats in the CSAT, the TPATs of local high schools predominantly focus on testing lexical knowledge and reading skills. However, this appears to be in conflict with the main goal of English language teaching in secondary schools in Korea, which is to develop students’ communicative competence for practical and academic purposes (Ministry of Education, Science, and Technology, 2011). If the TPATs are generally limited to receptive language skills (i.e., lexical knowledge and reading comprehension), developing students’ communicative competence would be difficult to achieve, as the relative effects of focusing predominantly on such test items on the acquisition of communicative competence remain unclear.

High-stakes test

The TPATs that are administered to high school students are regarded as high-stakes tests, since every test score is proportionally calculated into one of nine scales, which is an important criterion required for admission into universities in Korea. For example, if a student’s score in the TPAT is above the scores of the students in the top 11% of their high school but below the scores of the students in the top 4%, he will be graded at Scale 2 (see Table 7).

Table 7. Grade scales of students’ scores in achievement tests

Scale	1	2	3	4	5	6	7	8	9
Ratio	4%	7%	12%	17%	20%	17%	12%	7%	4%
Cumulative	4%	11%	23%	40%	60%	77%	89%	96%	100%

Most notably, the problem with this type of grading system is that there is a fundamental gap between the purpose of the test and the grading system used to assess the student's ability. Specifically, the TPATs are criterion-referenced, which purport to assess whether the student has acquired a certain level of ability or a domain of contents; however, the nine-scale grading system which is currently being used to rate the students is norm-referenced and thereby measures individual student's score in relation to other students' scores—not based on a specific standard or criterion. In other words, the teacher must classify each student into one of the nine scales, irrespective of the goals and learning outcomes of the class.

In order to differentiate among individual students, Korean high school teachers must consistently rely on various strategies in test development and administration. For example, an unnecessarily lengthy reading text containing three to four paragraphs is given as a stimulus for a single test item to make it difficult for the majority of the students to produce a correct answer under time pressure. Another example involves 25 pairs of English words with the Korean translation from which the students must select one incorrect pair.

Specifically, these types of challenging test items were observed most frequently in the test of School E, which is located in a school district well-known for its high education fever and competitiveness. Given this contextual challenge, the teachers at School E were particularly keen to develop exceedingly difficult tests in order to assure an anticipated number of students to systematically fit into each of the nine scales of the predetermined grading system. The following comment by a teacher explicitly reflects the extent to which the teachers strive to include difficult test items in the TPAT: "We try virtually everything to rate the students across the nine grading scales." In such a competitive academic environment, a student who has trouble solving absurdly difficult questions often becomes demoralized which in turn adversely impacts future learning outcomes across the entire curriculum. Although incorporating certain challenging test items may be necessary for rating students, the use of this type of sorting system may have a detrimental effect on the students' attitudes toward learning English and hence their performance in school.

VI. Conclusion

The investigation of the TPATs in the eight high schools undertaken in the present study has identified two sociopolitical and contextual factors—*the restrictions of the test format* and *a high priority given to college admissions test preparation*—which fundamentally dictate how individual test items are constructed. Administrative restrictions placed on TPATs have a great impact on the macrostructure as well as the microstructure of the test. As the TPATs have a range of limitations to testing venue, time, and scoring process, one of the most serious factors contributing to quality problems inherent in TPATs is the absence of free-production test items as an authentic measure of English language proficiency. A primary reason for a complete lack of free-production items in the eight TPATs investigated in this study is that the schools are required to administer each test in compliance with policy regulations according to which all students must take the test at the same time in the same place.

A high priority given to college admissions test preparation is an additional dominant factor that shapes the TPATs, which are designed to encourage students to focus not only on what is taught in class but also on other external study materials (e.g., vocabulary textbooks and mock CSAT) that are considered important in improving the CSAT score, which is the most crucial parameter in university admission. Along with the CSAT, the TPAT itself is a high-stakes test for university admission qualification as the students' scores in the TPAT are used as a two major criterion for college acceptance. The grading system of TPAT is, however, norm-referenced: students are rated according to their relative achievement level among the total students in the same grade at the school. This scheme has led the teachers to devise a variety of strategies for sorting students evenly into the established nine scales of the grading system. Most important, the TPATs must be designed to yield a very small number of students to be rated at the top scales of the grading system. As a natural consequence, some test items become unnecessarily difficult, obscuring the genuine purpose of achievement tests—assessing students' achievement and improving instructional effects (Popham, 2003).

One way to offer a more comprehensive English language education may constitute a dramatic revision in language testing practices. For example, testing methods need to incorporate a test of written English competence which measures the student's ability to compose a variety of written texts with a coherent organization. In an exam-oriented education system that still dominates the majority of EFL contexts, it may be the single most important element for improving the overall quality of English language education. In addition, the fundamental role of high school TPATs should not be solely to screen students for university admission. Given that a majority of Korean students pursue higher education, high school TPATs should be particularly utilized to promote interactional competence in English, which is increasingly becoming crucial for success in university. If the TPATs of English in high schools continue to be rife with obscure reading comprehension questions while excluding free-production items, students in this type of school system will be ill-prepared to meet the demands of college-level English.

Overall improvement of TPATs of English in high schools is possible only when proper attention is given to the comprehensive process of planning, administering, and using the CBA. As a multitude of sociopolitical and contextual factors impose restrictions on the TPAT, in-service training and specific assessment guidelines should be implemented in local schools. In addition, collaboration among the teachers is imperative to enhance the quality of the TPATs. As the best model of TPAT in the present study (School C)² suggests, the English teachers are recommended to initially develop a pool of 80 to 90 test items together, from which 25 to 30 of the most relevant questions should be selected and administered to the teachers themselves first as a starting point for discussion on the efficacy of each test-item. Furthermore, it is important for the teachers to engage in ongoing discussion about issues of assessment, particularly how to devise better

² As presented in Table 6, the smallest number of violations against evaluation criteria was observed in the tests of School C. In addition, the test items of School C were judged as well-written and meaning-based.

tests as a way of preparing high school students not merely for the CSAT but for college coursework and beyond.

The present study investigated the TPATs of English in eight Korean high schools and explored the quality as well as the relevance of actual tests that are developed by local schoolteachers. In order to generalize the findings to broader contexts, further research should conduct a nation-wide investigation of TPATs in diverse school settings. In addition, incorporating students' perceptual data using qualitative techniques would provide a new diagnostic insight into the testing and assessment practices of local high schools. The results could be a clearer understanding of specific organizational processes and practices that may help instill high quality TPATs while promoting appropriate management of sociopolitical and contextual factors which currently adversely affect the TPATs.

References

- Allen, P., Fröhlich, M., & Spada, N. (1983). *The communicative orientation of language teaching: An observation scheme* (Reports). Ontario: Ontario Institute for Studies in Education.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing, 15*(1), 45–85.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing, 18*(4), 393–407.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Byun, G. (2002). What authentic English tests should contain to improve the efficiency of English education. *Modern English Education, 3*(2), 135–148.
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39–62.
- Coniam, D. (2009). Investigating the quality of teacher-produced tests for EFL students and the effects of training in test development principles and practices on improving test quality. *System, 37*(2), 226–242.
- Coniam, D. (2014). Pursuing the qualities of a “good” test. *Frontiers of Education in China, 9*(2), 238–249.
- Clarke, S., & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996–1998. *Evaluation and Research in Education, 14*(1), 38–52.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305–334.
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching*.

- New York: Macmillan.
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill Book Company.
- Hill, K. (2012). *Classroom-Based assessment in the school foreign language classroom*. Frankfurt: Peter Lang.
- Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420.
- Im, B. (1995). Problems in high school English teacher-made tests. *English Teaching*, 50(1), 1–20.
- Im, B. (1996). A suggestion for better test construction in secondary school teacher-made tests. *English Language and Literature Teaching*, 2, 1–23.
- Jeon, B. M., Park, J. E., Ahn, B. K., Oh, J., Yu, J. M., Lee, S., & Kim, S. H. (2005). A study on improving English teacher's competence in student assessment. *English Teaching*, 60(2), 3–27.
- Jeong, H.-J. (2012). *Effects of students' feedback on English teachers' test preparation and administration*. Unpublished master's thesis. Seoul National University, South Korea.
- Jung, S., & Norton, B. (2002). Language planning in Korea: The new elementary English program. In J. Tollefson (Ed.), *Language policies in education: Critical issues* (pp. 245–265). Mahwah, NJ: Lawrence Erlbaum.
- Jung, C. S. (2007). A study on item analysis of the intramural regular test and on correlation between the intramural regular test and the national achievement test developed by KERINET. *The Journal of Curriculum and Instruction Studies*, 2(1), 149–185.
- Kim, S. A. (2009). An analysis of the multiple-choice test items constructed by middle school English teachers. *Korean Journal of Applied Linguistics*, 25(2), 143–169.
- Kim, Y.-H. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26(2), 187–217.
- Korea Institute for Curriculum and Evaluation. (March 28, 2012). A new release: Masterplan of 2013 college scholastic

- achievement test. Retrieved from <http://www.suneung.re.kr/board.do?boardConfigNo=57&page=8&menuNo=230&action=view&boardNo=12112>.
- Kwak, S. (2008). *Analysis of English test items in Korean high schools* [in Korean]. Unpublished master's thesis. Korea University, South Korea.
- Lado, R. (1970). *Language testing: The construction and use of foreign language tests*. London: Longman Group Ltd.
- Leung, C. (2005). Classroom teacher assessment of second language development. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 869–888). Mahwah, NJ: Lawrence Erlbaum.
- Leung, C. (2007). Dynamic assessment: Assessment for and as teaching? *Language Assessment Quarterly*, 4(3), 257–278.
- Na, I. (2012). *An action research on improving the process of school-based English assessment in middle school*. Unpublished master's thesis. Ewha Womans University, South Korea.
- Nitko, A. J. (1989). Designing tests that are integrated with instruction. In L. Robert (Ed), *Educational measurement: The American Council on Education/Macmillan series on higher education* (pp. 447–474). New York, England: Macmillan Publishing.
- Madson, H. S. (1983). *Techniques in testing*. New York: Oxford University Press.
- Ministry of Education, Science, and Technology. (2011). *Yeongeogwa gyoyukgwajeong [English curriculum]*. Seoul, Korea: Ministry of Education, Science, and Technology.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462.
- Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation1. *International Journal of Applied Linguistics*, 16(2), 163–188.
- Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and

pitfalls. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 505–520). New York: Springer US.

Rivers, W. M. (1981). *Teaching foreign language skills* (2nd Ed.). Chicago: The University of Chicago Press.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language testing*, 13(3), 298–317.

Shih, C.-M. (2010). The washback of the general English proficiency test on university policies: A Taiwan case study. *Language Assessment Quarterly*, 7(3), 234–254.

Authors

Sung, Min-Chang

Seoul National University, 1st author, corresponding author
potamin3@snu.ac.kr

Jo, Victoria H.

Seoul National University
vhjovhjo@snu.ac.kr

Appendix. Sample test items

Test items on reading skill. Test items on reading are intended to measure a variety of capacity involved in foreign language reading such as awareness of coherence/cohesion, identification of entities referred by pronouns, and understanding of topics and details. For example, the following item is designed to assess whether the test-taker is able to understand specific information in the given passage:

Read the passage and choose an incorrect statement.

Nicolette was filled with excitement. She contacted everyone she knew and asked them for their support. Progress was slow at first, but she refused to give up. Then she made contact with Oliver Wyss, a soccer coach she had met while doing volunteer work at a summer camp. Oliver is a professional soccer player and the founder of *Soccer for Hope*, an organization that helps children with severe illnesses. Oliver liked Nicolette's idea and agreed to help set up a bank account for *Futbol 4 Refugees* so that people could easily send donations. Oliver also donated the first 25 soccer balls.

- ① Nicolette and Oliver met each other at a summer camp.
- ② Oliver is a soccer coach as well as an establisher of *Soccer for Hope*.
- ③ Oliver agreed to support Nicolette's idea.
- ④ Oliver has cured many children with serious diseases.
- ⑤ Oliver is the first one who donated the soccer balls to *Futbol 4 Refugees*.

(Extracted from the test of School C)

Test items on writing skill. Assessment of writing can be conducted at a variety of levels ranging from word- or sentence-level writing to paragraph- or essay-level writing. In addition, students can be asked to freely express their opinions or be given pseudo-writing tasks such as scrambled sentences, fill-in-the-gap, or sentence-level translation. For example, the following item provides a Korean sentence and asks test-takers to translate it into an English sentence:

Translate the following Korean sentence into English.

Sinbiha-n geos-i sarm-eseo wedaeha-n geos-del-el kyeongheomha-nun de iss-eoseo kunbon-cheok yeokhak-ul ha-nda.

Translation: _____

(Extracted from the test of School H)

Test items on listening skill. Test items on listening, in general, play recorded speeches or conversations and assess how well test-takers understand the recordings. In the following example, students are asked to listen to a conversation about school foods and to find out an option with incorrect information:

Listen to the conversation and choose an incorrect statement.

- ① Children are on a diet.
- ② Children have health problems.
- ③ Children eat junk food at school.
- ④ Children eat fatty and sugary snacks.
- ⑤ The cafeteria food is not nutritious.

(Extracted from the test of School G)

Test items on speaking skill. Given that the TPAT is administered in a paper-based format, it is impossible to record a test-taker's spoken production. Therefore, indirect measurement of speaking skills has been adopted (e.g., selecting the most unnatural utterance in a conversation):

Choose the most unnatural utterance in the following conversation.

- A: It's already 7 o'clock. ① I'm running out of time. How can I finish the volunteer project my boss asked for?
- B: Take it easy. You have two days to go. ② Mind your own business. Why don't you postpone it till tomorrow?
- A: You must be kidding. I don't think that delaying the project is a good idea. The mistakes I made before are still bothering me.
- B: ③ Let bygones be bygones. Everything will be fine. Besides, taking your time can give you a hand to solve a problem sometimes.
- A: ④ I couldn't agree with you more. Let's get out of here and have dinner. It's my treat.
- B: I'd love to but, ⑤ can I take a rain check on that? I'm planning to go to an exhibition with my sister.

(Extracted from the test of School G)

Test items on grammatical knowledge. Items on grammar usually examines whether a test-taker can recognize ungrammatical phrases and/or correct them. The test stimulus may present several sentences or even paragraphs from which an ungrammatical phrase is to be selected:

Read the passage and choose an ungrammatical word/phrase.

Currency has existed for hundreds of years as a means of trading goods. In today's economy, many currencies - such as the won, yen, dollar, and euro - ① are exchanged every day. In addition to its value as money, currency has a certain symbolic value. The images ② shown on paper bills and coins reveal important political features. Thus, currency can teach us a lot about the world. There are more than 190 countries on six continents, all of ③ them are unique and interesting to visit. If we were rich, we ④ would be able to visit them all. However, without ⑤ spending any money, we can still take a trip all around the world.

(Extracted from the test of School G)

Test items on lexical knowledge. Items on vocabulary assess a wide range of lexical knowledge including definition of words, awareness of synonyms and antonyms, and use of contextually appropriate words:

Choose the word that has the closest meaning to the underlined word.

He pleaded for mercy in the courtroom.

- ① begged ② implied ③ presented ④ employed ⑤ processed

(Extracted from the test of School E)

Received: 2nd November, 2015

Reviewed: 29th November, 2015

Revised version accepted: 23rd December, 2015