

A Comparison of the Texts Used in the Secondary School English Tests

Sung-Ae Kim
Pusan National University

Kim, Sung-Ae. (2011). A comparison of the texts used in the secondary school English tests. *Modern English Education*, 12(4), 215-236.

The purpose of this study is to investigate the similarities and/or differences of the secondary school English tests in their ways of using the texts. For that purpose the study analyzed 30 middle school and 30 high school English tests, focusing on the number and types of the texts used and the kinds and proportion of the knowledge measured with the texts. In order to see the similarities and/or differences between the middle school and high school tests and among the tests for the same grade and for different grades the study compared the tests administered in different middle schools and high schools and the tests for different grades and for the same grade in each school level. According to the result of the study, there were both similarities and differences among the tests for different school levels and for different grades in their ways of using the texts. However, despite the common features found in the tests for the same grade, there was also a wide variation among the tests administered in different schools. Based on the result, the study concludes that due to the lack of the established criteria and thus the teachers' arbitrary use of the texts in constructing the tests, the secondary school English tests fail to maintain full consistency and are not considered effective enough for the development of the target language ability in the students.

[English` tests/secondary school English/test items/item construction/
영어시험/중등영어/평가항목/문항개발]

I. INTRODUCTION

Consistency has been argued as an important factor for efficient learning (Tuttle, 2009). Unfortunately, however, based on my own observations of the English classes from the elementary to high school level, I have a strong feeling that the English education in this country has not been very successful in providing the students with consistent education. In

the elementary school English classes, most of the class time is devoted to the development of the spoken English ability with various “communicative” activities (the term “communicative” here does not necessarily mean that the activities used in the class were actually communicative, though.), whereas such activities diminish in the middle school English classes, and by the time the students reach the high school level few, if any, classes virtually employ any communicative activities. Moreover, even in the same school individual teachers may not be the same in their choice of the teaching approach. Especially in the middle schools some teachers with sufficient English proficiency feel more confident in teaching their classes in English and are more willing to adopt the communicative approach. On the other hand, other teachers with insufficient English proficiency feel less confident and are reluctant about adopting the same approach. Such differences among individual teachers often result in inconsistency in the teaching approach through which their students learn the subject. For instance, it is possible that the same students learn by one approach chosen by the teacher(s) in the middle school and by another approach chosen by the teacher(s) in the high school. It is also possible that even in the same school teachers may have different views about an effective approach and thus practice their teaching following a different approach. From the students’ standpoint such inconsistency in the teaching approach can be confusing and it may result in reducing the effectiveness of their learning.

Motivated by the concern about the inconsistency that seems to exist in our English education, the present study aims to investigate how similar and/or different the secondary school English tests are from each other. The specific research questions addressed are 1) Are there any differences among the tests in the number and types of texts used? 2) Are there any differences among the tests in the kinds and proportion of knowledge measured with the texts? It is expected that the result of this study would reveal useful information about how consistent or inconsistent the secondary school English tests are in the use of the texts and about what is necessary to bring about the beneficial backwash with the tests.

II. BACKGROUND

There have been arguments regarding the importance of the texts used in the tests. Bailey (1998), for instance, argues that texts in the language test are used as the stimulus materials to measure the test taker’s language ability. As such, he claims, they must be able to allow the teacher to elicit an appropriate sample of behavior from the test-taker. However, without a deep consideration into various factors including the skills to be measured, there is a danger that errors may be introduced in the test. According to Rudner and Schafer (2002), such errors may be introduced in selecting the tests for a test

and as the skills become more complex, more errors are likely to be introduced. In fact, both the selection of the texts and the way they are used in the test can make a difference in the focus as well as difficulty of the test and hence, overall quality of the test. Given the importance of the texts used in the test it goes without saying that in order to make the test bring about the beneficial backwash the test constructor must possess the necessary knowledge and skills in the selection and use of the texts and thus make the test valid enough.

Generally speaking, texts are classified into various types on the basis of the intent of the communicator (Larson, 1984). For instance, the communicator's intent of a narrative text is to tell about a sequence of events, that of an expository text is to explain something, and that of a descriptive text is to list characteristics of something. Regarding text typology, many specific types have been recognized since Aristotle's day, but consensus seldom exists in typological analyses and the resulting text typologies have some shared as well as some unshared categories (Beghtol, 2000; Fludernik, 2000). For instance, in Kinneavy's (1971) analysis there are expressive, referential, literary and persuasive types of text, each of which again splits into subcategories. In Longacre's (1983) analysis there are narration, procedural discourse, behavioral discourse, and expository discourse. On the other hand, Chatman (1990) differentiates among descriptive, argumentative, and narrative types of text.

Despite the various ways of categorizing the texts, however, it is certain that sorting text types on the basis of the communicative intents poses special problems. This is because each text can contain elements of more than one purpose (Beghtol, 2000). Thus, even the most familiar text types are, Beghtol claims, unstable, changeable and can divide, fuse and/or mutate to form different kinds of hybrid texts, for which new names, such as non-fiction novel, informercial, prose poem and docudrama, are often coined. With such a problem involved in sorting text types, research studies on text types have suggested that different text types involve different processing mechanisms (Cote, Goldman & Saul, 1998; Trabasso & Magliano, 1996). In order to help the students gain control over different types of text in the real world, therefore, it will be necessary to have them become familiar with a variety of text types in their English learning experience in school.

The important role of the texts in constructing the tests suggests that the freedom that the teachers have in selecting and using the texts can make a big difference in deciding the focus of the test and thus directing the students' study behaviors (Sung-Ae Kim, 2010a). For instance, if the texts used in the test are mostly conversation type, the test would encourage students to focus more on the conversation skills. On the other hand, if the texts used are mostly description type, the test would encourage the students to focus more on the reading skills.

In constructing the test what kinds of knowledge are measured with the texts or what tasks are required with the texts is no less important than the types of texts used. This is because the kinds of knowledge measured or the task required in the tests inevitably influence the test taker's behavior (Anderson, Bachman, Perkins & Cohen, 1991; Bachman, 1990; Bernhardt, 1991; Gordon & Hanauer, 1995). If the test items made with the texts mostly measure the linguistic (i.e., vocabulary and grammar) knowledge rather than the content knowledge, for instance, then the students will focus on studying vocabulary and grammar rather than on comprehending the content of the text. However, despite the importance of the task required or the knowledge measured in the test, it seems true that not all the tests are successful in this respect.

Given the importance of what types of texts are used and what kinds of knowledge are measured with the texts in constructing the test, the test constructor must possess the necessary knowledge and skills to select the texts and use the texts effectively enough to construct a valid test. However, without any criteria and guideline provided on how to use the texts in constructing the tests (Sung-Ae Kim, 2010a, 2010b), the secondary school English teachers mostly rely on their own discretion in using the texts. It would be natural, therefore, to assume that despite the common features in the tests due to the influence of the textbooks following the national standard imposed by the Ministry of Education, Science and Technology, there could also be differences in individual teachers' ways of using the texts in test construction. Recognizing the variations among the tests administered in different schools, studies within the country have recently proposed the establishment of professional standards for student assessment (Byoung-Man Jeon, Jun-Eon Park, Byung-kyoo Ahn, Junil Oh, Jay Myong Yu, Soyoung Lee & Shin-hye Kim, 2005; Sung-Ae Kim, 2009, 2010a, 2010b).

Teaching and testing must be intertwined in order to bring about the maximal effect of the two (Rudner & Schafer, 2002) and the information provided by the tests should be the tool that guides the teaching process. This implies that the teachers themselves should be able to assess validly and reliably the ability they want to foster in their students. The problem, however, is that constructing a valid and reliable test to measure such an ability is never easy. According to Shohamy (1992), this difficulty, at least partially, is responsible for lack of discussion on school testing in the area of language testing theory, research, and development. Kehoe (2002) rightly points out that such lack of discussion in turn has resulted in a notable concern of many teachers who must construct the tests without enough training or information to rely on.

Well-constructed tests not only assess student ability but also reveal a great deal about language instruction and thus can help the teachers make their teaching more effective (Liskin-Gasparo, 1984). Given the important role of the test, the job of the professionals would be to ensure that the tests administered in the schools are constructed in such a

way as to bring about the maximal effects on the learning process. Unfortunately, however, it must be admitted that the professionals in this country have not paid enough attention to what the teachers do in constructing the tests and as a result, it is true that we do not have enough information about the quality of the tests constructed by the teachers. Considering that tests must reflect what has been taught and that tests have a strong backwash effect on the students' learning behavior (Heaton, 1990; Hughes, 1989; Prodromou, 1995) and that texts play a crucial role in determining the quality of the tests, analyses of how the texts are used in the tests would be quite meaningful.

III. METHOD

1. Data Collection

The data for the analysis were collected from a web site (zocbo.com) which gives out information for fees about the tests already administered in various schools. All the tests posted on the site are the scanned versions of the original tests and there can be no room to change any part of the original test administered in the selected school either in the form or content. For the purpose of the study 30 English tests from six middle schools and 30 English tests from six high schools were randomly selected from each of the five school districts in a metropolitan city. Among the 30 tests thus collected for each school level, there were ten tests for each grade level. Even though the data collected were far from being exhaustive, ten tests for each grade level were considered to be sufficient to gain information about how texts are used in middle school and high school English tests. A summary of the information about the data is given in Table 1.

Table 1
Information about the Data

School District	School Level	# of Schools	Test - Grade						Total # of Tests
A	M	6	MA1-1,	MA2-1,	MA3-2,	MA4-2,	MA5-3,	MA6-3	60
	H	6	HA1-1,	HA2-1,	HA3-2,	HA4-2,	HA5-3,	HA6-3	
B	M	6	MB1-1,	MB2-1,	MB3-2,	MB4-2,	MB5-3,	MB6-3	
	H	6	HB1-1,	HB2-1,	HB3-2,	HB4-2,	HB5-3,	HB6-3	
C	M	6	MC1-1,	MC2-1,	MC3-2,	MC4-2,	MC5-3,	MC6-3	
	H	6	HC1-1,	HC2-1,	HC3-2,	HC4-2,	HC5-3,	HC6-3	
D	M	6	MD1-1,	MD2-1,	MD3-2,	MD4-2,	MD5-3,	MD6-3	
	H	6	HD1-1,	HD2-1,	HD3-2,	HD4-2,	HD5-3,	HD6-3	
E	M	6	ME1-1,	ME2-1,	ME3-2,	ME4-2,	ME5-3,	ME6-3	
	H	6	HE1-1,	HE2-1,	HE3-2,	HE4-2,	HE5-3,	HE6-3	

M: Middle School, H: High School

Even though there may be other types of tests also administered in each school, such as the listening test or the performance test, the mid-term and final exam administered in each semester, are known to be the main type of test across all the schools. As such, analysis of how texts are used in those tests is expected to reveal useful information about the focus of the tests and about what is required from the students.

2. Data Analysis

The analysis focused on the number and types of the texts used and the kinds of knowledge measured with the texts. The number of the texts used in each test was calculated simply by counting the frequency of them. What was counted as a text in the analysis was two or more related sentences which were used as the stimulating material in constructing the test item(s). In order to investigate what types of texts were used in the tests, each text was analyzed and described for the type. The resulting types of text from the analysis were then divided into some major categories. Since the categories used to describe the types were grounded on the data themselves, the approach is considered to be inductive in nature. It must be admitted, though, that the categorization was inevitably affected by some arbitrariness due to the overlapping features of some categories and also due to the need to simplify the too many categories otherwise. However, such arbitrariness was not expected to seriously affect the overall interpretation of the data.

The kinds of knowledge measured with each text were categorized into the linguistic knowledge (vocabulary & grammar knowledge), content knowledge, and other knowledge. The proportion of each kind of knowledge was calculated by the percentage of the test items made on that knowledge. When the test taker can choose the correct answer with the vocabulary knowledge and/or the grammar knowledge only, the knowledge measured was categorized as linguistic knowledge. On the other hand, when the test taker must also understand the content of the text, not just the vocabulary and/or grammar, in order to choose the correct answer, the knowledge measured was categorized as content knowledge. When the knowledge measured was neither linguistic knowledge nor content knowledge, it was categorized as other Knowledge. The test items that measure other knowledge include those items which measure the students' knowledge of punctuation, common sense, etc. Since the categories emerging from this analysis are grounded on the data themselves, the approach is considered to be inductive.

In order to see how similar and/or different the tests are between the middle school and the high school, the study compared the middle school tests and the high school tests in the number and types of the texts used and the kinds of knowledge measured with the texts. In addition, in order to see if there are any similarities and/or differences among the tests for the three grades in each school level (middle school and high school), the study compared

the tests among the grades in each school level. In order to see if there are variations among the tests for the same grade, the study also compared the tests for each grade which were administered in different schools. These comparisons were expected to reveal how consistent the English tests are between the middle school and the high school, among the grades within the same school level, and among the same grade in different schools.

To measure the significance of the difference among the tests, a statistical analysis using a one-way ANOVA with post-hoc multiple comparison test using Tukey HSD was administered. The significance level was set at $p < .05$ for the statistical decision. In order to assure the reliability of the analysis, two coders, the researcher herself and a trained graduate student, independently worked on the data and the results of each coder's analysis were compared. The inter-coder reliability at the first coding was .84. However, when there were differences between the two results of the coding, the differences were eliminated through discussion and recoding. Even though the information obtained from the analysis is far from being exhaustive, focusing on only some limited aspects of the texts used in the tests, it is expected to provide meaningful information about how consistent the secondary school English tests are and about what is necessary to bring about the beneficial backwash in our English education.

IV. RESULTS AND DISCUSSION

1. Number and Types of Texts Used

As can be noticed in Table 2, one big difference in the use of the texts in the middle school and high school English tests is that the number of the texts used in the test is much bigger in the high school tests than in the middle school tests. The mean numbers of the texts used in the tests for the 1st, 2nd and 3rd grade in the middle school were 10.3, 9.5, and 12.9, respectively ($p = .033^*$), whereas those for the 1st, 2nd and 3rd grade in the high school were 22.1, 24.4, and 28.6, respectively ($p = 0.001^*$), showing a statistical significance among the means. However, a post-hoc multiple comparison test using Tukey HSD for the mean number of texts used in the tests shows that in both the middle school and high school tests there is a statistically significant difference in the use of texts between the 2nd grade and the 3rd grade, but not between the 1st grade and the 2nd grade (See Tables 3 & 4). This indicates that in both the middle school and high school tests, more weight is put on the texts in the tests for the 3rd grade than for the lower grades, suggesting that there is probably some gap in the focus of the teaching between the 3rd grade and the lower grades.

Table 2

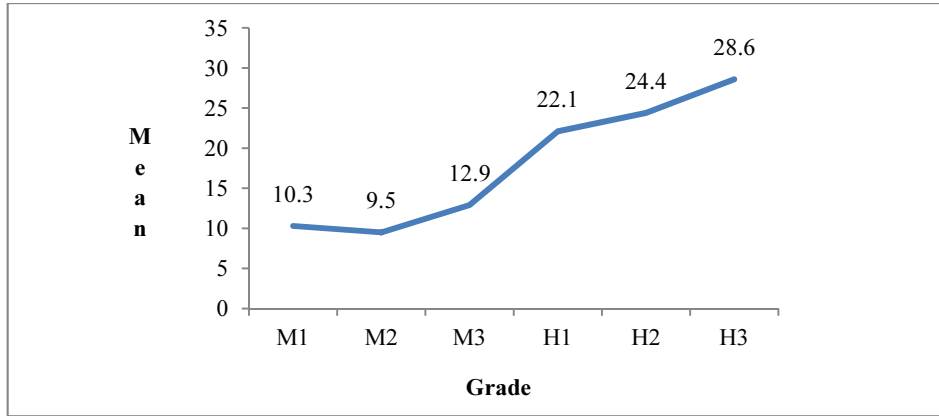
Number and Types of Texts Used in the Tests

School Level	Grade	Number (%) and Types of Texts Used					
		Conv.	Letter	Lyric	Des.	Adv.	Mean
Middle School	1st (10)	50 (48.54%)	8 (7.77%)	1 (0.97%)	42 (40.78%)	2 (1.94%)	10.3 (100%)
	2nd (10)	23 (24.21%)	5 (5.26%)	0 (0%)	64 (67.37%)	3 (3.16%)	9.5 (100%)
	3rd (10)	34 (26.36%)	5 (3.88%)	1 (0.76%)	89 (68.99%)	0 (0%)	12.9 (100%)
	p	0.017*	0.722	0.612	0.000*	0.452	0.033*
High School	1st (10)	11 (4.98%)	2 (0.90%)	0 (0%)	207 (93.66%)	1 (0.45%)	22.1 (100%)
	2nd (10)	6 (2.46%)	0 (0%)	0 (0%)	238 (97.54%)	0 (0%)	24.4 (100%)
	3rd (10)	1 (0.35%)	0 (0%)	0 (0%)	284 (99.30%)	1 (0.35%)	28.6 (100%)
	p	0.158	0.125	0	0.000*	0.612	0.001*

Conv.: Conversation, Des.: Description, Adv.: Advertisement

As can be noticed in the table, one interesting finding is that there is a sudden leap between the 3rd grade in middle school and the 1st grade in high school in the number of the texts used. This seems to reflect that more weight is put on the text in the high school tests than in the middle school tests. In fact, the researcher could also notice in the analysis that most of the test items in the high school English tests are made with a text, whereas it is not always the case in the middle school English tests. A graphic display of the mean number of the texts used in the tests for each grade is provided in Figure 1.

Despite the general features mentioned above, however, a closer look at the data reveals that there also exists a great variation among the tests even for the same grade. In the tests for the 1st grade in the middle school the number of the texts used ranged from 3 (MD1-1) to 15 (MB1-1). Similarly, in the tests for the 2nd grade it ranged from 7 (MA4-2) to 13 (MD4-2). Also, in the tests for the 3rd grade it ranged from 9 (MD5-3) to 17 (MD6-3) (See Appendix A for details).



M: Middle School, H: High School

Figure 1. Mean Number of Texts Used in the Tests for Each Grade.

Based on the result of the analysis of all the secondary tests in terms of the types of text used, the types of text used in the secondary school English tests can be divided into five categories: Conversation, Letter, Lyric, Description, and Advertisement. Here, Description includes those texts which describe about people, things, or events. Despite the five categories emerging from the analysis, however, the types of text actually used in the tests were very limited. As shown in Table 2 above, the two most frequently used types across all the three grade levels in the middle school were Conversation and Description and the rest three types (Letter, Lyric, and Advertisement) were used only rarely (See Appendix A for the details of individual tests). In the case of the high school tests, on the other hand, most of the texts were Description. That is, 93.66% of the texts used in the tests for the 1st grade, 97.54% for the 2nd grade, and 99.30% for the 3rd grade were Description.

Table 3

Result of Post-hoc Test for the Mean Number of Texts Used in MS Tests

	(I) Grade	(J) Grade	(I-J) Mean Difference	Std. Err.	p
Tukey HSD	1	2	.80000	1.27512	.806
		3	-2.60000	1.27512	.122
	2	1	-.80000	1.27512	.806
		3	-3.40000	1.27512	.033*
	3	1	2.60000	1.27512	.122
		2	3.40000	1.27512	.033*

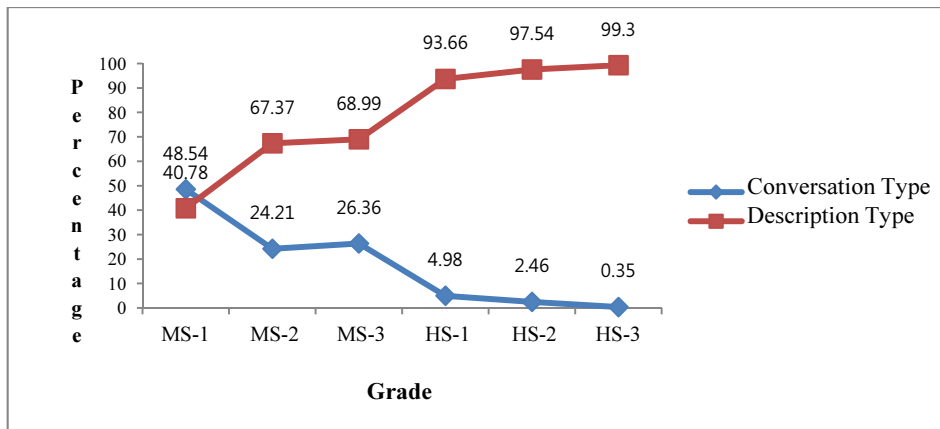
Table 4

Result of Post-hoc Test for the Mean Number of Texts Used in HS Tests

	(I) Grade	(J) Grade	(I-J) Mean Difference	Std. Err.	p
Tukey HSD	1	2	-2.30000	1.46995	.278
		3	-6.50000	1.46995	.000*
	2	1	2.30000	1.46995	.278
		3	-4.20000	1.46995	.021*
	3	1	6.50000	1.46995	.000*
		2	4.20000	1.46995	.021*

An ANOVA analysis with post-hoc multiple comparison test using Tukey HSD also reveals that in the middle school there is a sudden change between the tests for the 1st grade and the 2nd grade in the types of text used. As shown in Table 2 above, the most frequently used type of text in the tests for the 1st grade in the middle school was Conversation, whereas that for the 2nd and 3rd grade was Description. More specifically, in the tests for the 1st grade 48.54% of the texts used were Conversation, whereas in the tests for the 2nd and 3rd grade 24.21% and 26.36% of the texts were Conversation, respectively ($p=0.017^*$). Also, while 40.78% of the texts used in the tests for the 1st grade was Description, 67.37% and 68.99% of the texts used in the tests for the 2nd grade and the 3rd grade were Description, respectively ($p=0.000^*$). One interpretation of such a clear distinction between the 1st grade and the 2nd grade can be that for the 1st grade more focus is put on the speaking skills, while for the 2nd grade more focus is put on the reading skills.

Such a distinction found in the middle school tests between the Conversation type and the Description type, however, disappears in the high school tests. As the high percentages across all the three grades (93.66% for the 1st grade, 97.54% for the 2nd grade, and 99.30% for the 3rd grade) show, most of the texts used in the high school English tests are Description type (See Table 2), reflecting that the high school English test focuses on the reading skills. This is not a surprising result considering the known fact that despite the national English education policy encouraging the development of communicative language ability in the students, our high school English classes mostly focus on reading. A graphic display of the commonly used types of text in the secondary school tests is provided in Figure 2.



MS: Middle School, HS: High School

Figure 2. Types of Text Most Commonly Used in Secondary School English Tests.

Even though the analysis shows that in the middle school English tests Conversation type was the most frequently used type of text for the 1st grade and Description for the 2nd and 3rd grade, a closer look at the data reveals that there is also a wide variation among the tests even for the same grade. For instance, in the case of the Conversation type of text, one test (MB1-1) for the 1st grade used 10 texts, whereas other tests (MB2-1 and MD1-1) for the same grade used only 2 texts. Also, one test (MA5-3) for the 3rd grade used only 1 text, whereas another test (MD6-3) for the same grade used 7 texts. In addition, in the case of the Description type of text, one test (MA2-1) for the 1st grade used 7 texts, whereas another test (MD1-1) for the same grade used no text. Also, one test (MA6-3) for the 3rd grade used 14 texts, whereas another test (MD5-3) for the same grade used 5 texts (See Appendix A for details).

Such a wide variation among the tests also exists in the high school tests. For instance, even though most of the texts used in the high school tests are Description type, one test (HA1-1) for the 1st grade used 13 texts, whereas another test (HA2-1) used 27 texts. Likewise, one test (HA4-2) for the 2nd grade used 29 texts, whereas another test (HB4-2) for the same grade used 17 texts. The variation is reduced but there still exists some variation in the tests for the 3rd grade, too. One test (HC5-3) used 24 texts, whereas another test (HE6-3) used 31 texts. Such a difference in the number of the texts may suggest a difference among the tests in the lengths of the texts used, which in turn may be related to the difficulty of the test.

In summary, analysis of the data shows that there is a wide variation among the secondary school tests in the number and types of texts used. This variation may suggest that despite the common features among the tests, there are differences in the focus and probably the quality of individual tests administered in different schools, which is

suspected to result from lack of the established criteria and thus from the teacher's arbitrary use of the texts in constructing the test.

2. Kinds and Proportion of Knowledge Measured with the Texts

The two main categories of knowledge that emerged from the analysis of kinds of knowledge were linguistic knowledge and content knowledge. There was only one exception (i.e., one item in MA2-1) which measured the test taker's knowledge on punctuation and this was categorized as other knowledge (See Appendix C).

As shown in Table 5, despite the emphasis on meaning (or content) in the era of the communicative language teaching, a large part of the test items still measure the discrete point of linguistic knowledge in the secondary school English tests. In the middle school tests the mean percentage of the linguistic knowledge measured with the texts was 53.23% for the 1st grade, 53.42% for the 2nd grade, and 41.35% for the 3rd grade ($p=.375$). Similarly, in the high school tests the mean percentage of the linguistic knowledge measured was 43.66% for the 1st grade, 39.01% for the 2nd grade and 48.37 for the 3rd grade ($p=.173$).

Table 5

Proportion of Linguistic Knowledge and Content Knowledge Measured with the Texts

School Level	Grade	Mean # of Test Items Made	Mean % of Ling. K. Measured	Mean % of Cont. K. Measured	Total %
M.S.	1st	16.8	53.23	46.21	99.44
	2nd	18.6	53.42	46.58	100.0
	3rd	20.9	41.35	58.61	99.96
p			0.375	0.353	
H.S.	1st	26.0	43.66	56.54	100.0
	2nd	28.2	39.01	60.99	100.0
	3rd	30.6	48.37	51.36	100.0
p			0.173	0.173	

H.S.: High School, M.S.: Middle School, Ling. K.: Linguistic Knowledge, Cont. K.: Content Knowledge

As the result of the ANOVA analysis shows, there is no statistically significant difference among the grades either in the middle school or in the high school. However, a closer look at the data reveals that there is a wide variation within the same grade (See Appendices C & D for details.). For instance, in the middle school tests the variation in proportion of the linguistic knowledge measured ranged from 26.67% (ME2-1) to 100%

(MD1-1 & MD2-1) in the test for the 1st grade, from 23.08% (ME3-2) to 100% (MD4-2) in those for the 2nd grade, and from 9.17% (MC6-3) to 63.16% (ME5-3) in those for the 3rd grade. Similarly, the variation in the proportion of the content knowledge measured ranged from 0% (MD1-1 & MD2-1) to 73.33% (ME2-1) in the tests for the 1st grade, from 0% (MD4-2) to 63.16% (MD3-2) in those for the 2nd grade, and from 36.84% (ME5-3) to 70.83% (MC6-3) in those for the 3rd grade.

The same phenomenon can also be found in the high school tests. In the high school tests the variation in the proportion of the linguistic knowledge measured ranged from 26.09% (HD1-1) to 57.59% (HD2-1) in the test for the 1st grade, from 20.0% (HB4-2) to 48.28% (HC4-2) in those for the 2nd grade, and from 31.03% (HC5-3) to 77.42% (HB6-3) in those for the 3rd grade. Likewise, the variation in the proportion of the content knowledge measured ranged from 0% (MD1-1 & MD2-1) to 73.33% (ME2-1) in the tests for the 1st grade, from 0% (MD4-2) to 63.16% (MD3-2) in those for the 2nd grade, and from 36.84% (ME5-3) to 70.83% (MC6-3) in those for the 3rd grade.

Such a wide variation among the tests for the same grade may reflect that the focus of the test and hence, the quality of the test for the same grade administered in different schools may be different from school to school. If that is the case, it is not a problem to be ignored because it means there are differences in the ways the teachers in different schools measure their students' English ability. In fact, a glance of the test items made with the texts also revealed that they measured the students' discrete linguistic knowledge only and the students did not even have to read and understand the text to find the right answer. This shows that those teachers did not quite know why they used the texts for the test items, reflecting their lack of knowledge and skills in test construction.

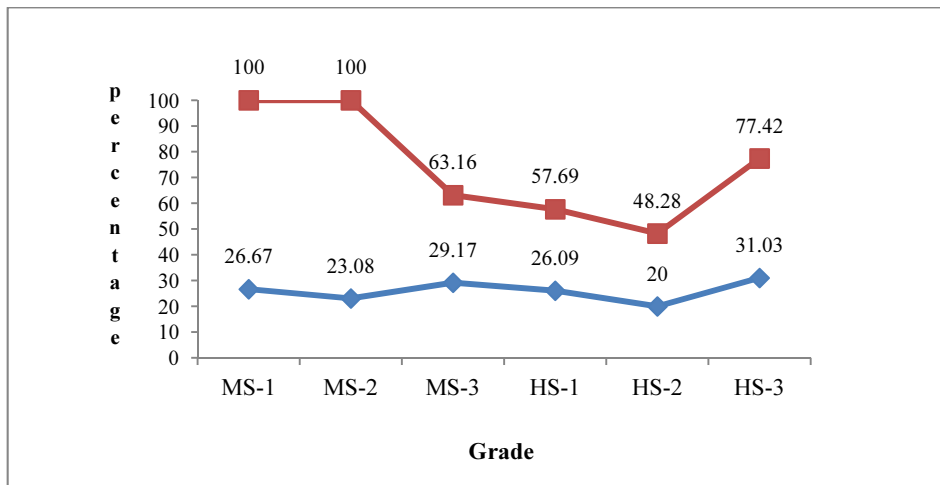


Figure 3. The Gap between the Highest and Lowest Proportion of Linguistic K. Measured.

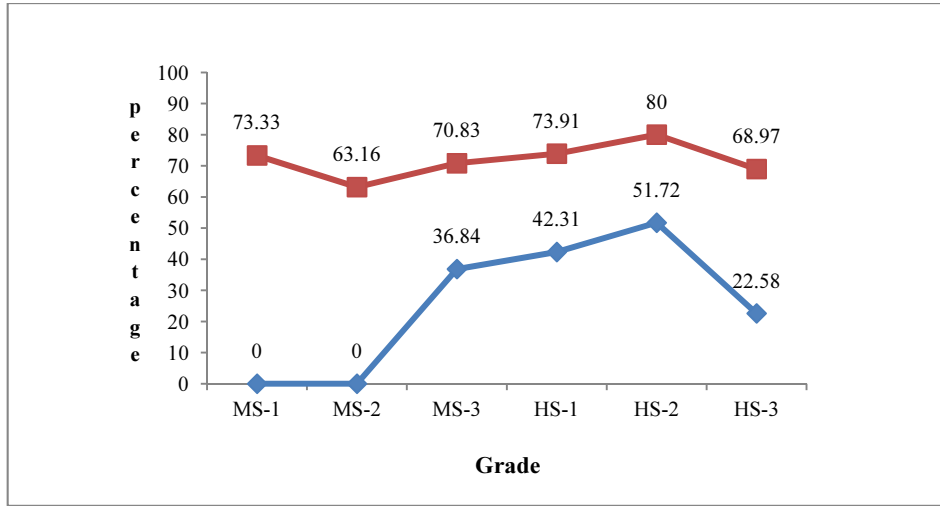


Figure 4. The Gap between the Highest and Lowest Proportion of Content K. Measured.

A graphic display of the highest and the low proportion of the linguistic knowledge and the content knowledge measured in the tests for the same grade is provided in Figures 3 and 4 above. As can be noticed, the gap is wider in the tests for the 1st grade and the 2nd grade in the middle school. Such a wider gap in the tests for the middle school 1st and 2nd grades may reflect the reality that without the pressure imposed by the College Scholastic Ability Test the middle school English teachers who teach the 1st and the 2nd grade have more freedom in deciding what to teach and how to teach than those who teach the higher grades.

In summary, analysis of the data shows that the kinds of knowledge measured with the texts were mostly limited to measuring the linguistic knowledge (i.e., vocabulary and grammar knowledge) and there is a wide variation among the tests even for the same grade. Analysis also shows that there is no statistically significant difference among the tests for different grades either in the middle school or the high school English in the proportion of the test items on either the linguistic knowledge or the content. This result seems to reflect the reality that despite the educational policy and efforts to implement the communicative language teaching and despite some teachers' efforts to change their practices, the focus of English teaching and testing in the secondary schools has not changed very much. In addition, the result also seems to reflect the reality that with lack of criteria and without the necessary training provided on how to make the test items with the reading texts, teachers are mostly on their own, which often results in a wide variation among the tests administered in different schools.

V. CONCLUSION AND SUGGESTIONS

Based on the results of the analyses, the study concludes that even though the tests administered in different schools share common features in their ways of using the texts, there also exists a wide variation among the tests administered in different schools even for the same grade. Individual tests show variations in the number and types of the texts used and also in the kinds and proportion of knowledge measured with the texts. It is suspected that the wide variation among the tests reflects lack of the established criteria and the necessary guideline and hence, the teachers' arbitrary use of the texts in test construction. Considering that the texts in the tests, as the stimulating materials, are crucial in deciding the quality of the test, it is necessary to establish a national standard to reduce the variations and keep an equal focus and quality among the tests and thus provide the students with consistent education. We need to remember that we can increase learning best when we are consistent in what we want the students to learn (Tuttle, 2009).

Described in the following are some suggestions that the researcher would like to make to enhance the consistency in our education and to make the tests bring about more beneficial backwash:

Firstly, there is a wide variation among the tests, which results from the teachers' arbitrary use of the texts and reflects the differences in the focus and quality of the tests. In order to reduce the variation, it is necessary to establish a national standard on how to use the texts in test construction.

Secondly, the texts used in the secondary school English tests at present are quite limited in the types. Considering that the EFL students have a limited source of target language input outside the classroom, it is necessary to allow the text materials to be included in the test that are not in the textbook but are used in class, so that it could induce the students to expand their experience with the target language. This, however, requires the decision and efforts of the professionals including policy makers and materials developers.

Thirdly, in order to encourage the students to become more familiar with as many text materials as possible through the study inside and outside the classroom, the restriction that is imposed on the middle school English tests by the educational policy that the content of the test must be based on the textbook needs to be modified. One cannot expect the students to pay much attention to the materials not covered in the tests. Teachers must be allowed to deal with more text materials in their testing.

Finally, it is necessary to develop various types of text materials for different proficiency levels. For now, our students simply do not have enough text materials to work on. They need to be provided with more text materials other than the textbook, which are leveled or placed in a sequence of difficulty progressing from simple to more complex (Cooper,

2001), and this can be done using different sets of criteria depending on the age and proficiency level being considered (Cooper & Kiger, 2001).

Given the claim that even formally and professionally developed tests often lack passage dependency (Bernhardt, 1991), it is not unexpected that the teachers, who have not received much help and guide in their in-service as well as pre-service programs, lack expertise in the construction of test items. It is argued that such lack of expertise among the teachers reflects lack of concern among the professionals in the field, as evidenced by lack of studies on this issue. More efforts are called for in the part of the professionals to solve this problem.

REFERENCES

- Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data scores. *Language Testing*, 8, 41-66.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decision, and directions*. New York, NY: Heinle & Heinle Publishers.
- Beghtol, C. (2000). The concept of genre and its characteristics. *Bulletin board of American society for information science & technology*. Retrieved August 26, 2009, from <http://findarticles.com/p/articles/miqa3991/is200012/ain8908335/>.
- Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical, and classroom perspectives*. Norwood, NJ: Ablex.
- Chatman, S. (1990). *Coming to terms: The rhetoric of narrative in fiction and film*. Ithaca, NY: Cornell University Press.
- Cooper, J. D. (2001). *Using different types of texts for effective reading instruction*. Houghton Mifflin Company. Retrieved August 26, 2009, from <http://www.eduplace.com/state/author/jdcooper.pdf>.
- Cooper, J. D., & Kiger, N. D. (2001). *Literacy assessment: Helping teachers plan instruction*. Boston: Houghton Mifflin Company.
- Cote, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relationship between processing and representation. *Discourse Processes*, 25, 1-53.
- Fludernik, M. (2000). Genres, text types, or discourse modes? Narrative modalities and generic categorization. Summer, 2000 issue of *Style*. Retrieved August 26, 2009, from <http://findarticles.com/p/articles/mim2342/is234/ai68279076/>.

- Gordon, C. M., & Hanauer, D. (1995). The interaction between task and meaning construction in EFL reading comprehension tests. *TESOL Quarterly*, 29(2), 299-324.
- Heaton, J. D. (1990). *Writing English language tests*. London: Longman.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jeon, Byoung-Man, Park, Jun-Eon, Ahn, Byung-kyoo, Oh, Junil, Yu, Jay Myong, Lee, Soyoun, & Kim, Shin-hye. (2005). A study on improving English teachers' competence in student assessment. *English Teaching*, 60(2), 3-27.
- Kehoe, J. (2002). Writing multiple-choice test items. In L. M. Rudner & W. D. Schafer (Eds.), *What teachers need to know about assessment* (pp. 69-74). National Education Association of the United States.
- Kim, Sung-Ae. (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, Sung-Ae. (2010a). An investigation of how texts are used in middle school English tests. *English Language Teaching*, 22(2), 157-178.
- Kim, Sung-Ae. (2010b). An analysis of the test items on text structures in high school English tests. *Korean Journal of Applied Linguistics*, 26(4), 251-274.
- Kinneavy, J. L. (1971). *A theory of discourse: The aims of discourse*. Englewood Cliffs, NJ: Prentice-Hall.
- Larson, M. L. (1984). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Liskin-Gasparro, J. E. (1984). Practical considerations in receptive skills testing. *Foreign Language Annals*, 17(4), 369-373.
- Longacre, R. E. (1983). *The grammar of discourse*. New York, NY: Plenum Press.
- Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, 49(1), 13-25.
- Rudner, L. M., & Schafer, W. D. (2002). Testing and teaching. In L. M. Rudner & W. D. Schafer (Eds.), *What teachers need to know about Assessment* (pp. 3-5). National Education Association of the United States.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
- Trabasso, T., & Magliano, J. (1996). Conscious understanding during comprehension. *Discourse Processes*, 21, 225-286.
- Tuttle, H. G. (2009). *Formative assessment: Responding to students*. Larchmont, NY: Eye on Education.

APPENDIX A

Number and Types of Texts Used in Middle School English Tests

Grade Level (# of Tests)	Test	Number & Types of Texts Used					Total	TOTAL (Mean)	P
		Con.	Let.	Ly.	Des.	Adv.			
1st (10)	MA1-1	6	1	0	3	0	10	103 (10.3)	0.033*
	MA2-1	5	1	1	7	0	14		
	MB1-1	10	0	0	5	0	15		
	MB2-1	2	0	0	5	0	7		
	MC1-1	7	0	0	5	0	12		
	MC2-1	7	0	0	4	0	11		
	MD1-1	2	1	0	0	0	3		
	MD2-1	3	0	0	6	0	9		
	ME1-1	4	1	0	3	0	8		
ME2-1	4	4	0	4	2	14			
2nd (10)	MA3-2	2	2	0	7	0	11	95 (9.5)	
	MA4-2	1	1	0	5	0	7		
	MB3-2	4	0	0	6	0	10		
	MB4-2	0	1	0	7	0	8		
	MC3-2	4	0	0	6	0	10		
	MC4-2	3	0	0	7	0	10		
	MD3-2	2	0	0	6	1	9		
	MD4-2	3	1	0	7	2	13		
	ME3-2	1	0	0	8	0	9		
ME4-2	3	0	0	5	0	8			
3rd (10)	MA5-3	1	0	0	9	0	10	129 (12.9)	
	MA6-3	2	0	0	14	0	16		
	MB5-3	3	2	0	9	0	14		
	MB6-3	3	0	0	8	0	11		
	MC5-3	5	0	0	10	0	15		
	MC6-3	2	0	0	12	0	14		
	MD5-3	2	2	0	5	0	9		
	MD6-3	7	0	1	9	0	17		
	ME5-3	4	0	0	6	0	10		
ME6-3	5	1	0	7	0	13			

APPENDIX B

Number and Types of Texts Used in High School English Tests

Grade Level (# of Tests)	Test	Number & Types of Texts Used					Total	TOTAL (Mean)	P
		Con.	Let.	Ly.	Des.	Adv.			
1st (10)	HA1-1	1	0	0	13	1	15	221 (22.1)	0.001*
	HA2-1	2	0	0	27	0	29		
	HB1-1	0	0	0	21	0	21		
	HB2-1	0	0	0	23	0	23		
	HC1-1	4	0	0	20	0	24		
	HC2-1	1	0	0	19	0	20		
	HD1-1	0	1	0	18	0	19		
	HD2-1	0	1	0	25	0	26		
	HE1-1	3	0	0	15	0	18		
	HE2-1	0	0	0	26	0	26		
2nd (10)	HA3-2	0	0	0	24	0	24	244 (24.4)	
	HA4-2	0	0	0	29	0	29		
	HB3-2	0	0	0	26	0	26		
	HB4-2	3	0	0	17	0	20		
	HC3-2	3	0	0	22	0	25		
	HC4-2	0	0	0	25	0	25		
	HD3-2	0	0	0	21	0	21		
	HD4-2	0	0	0	21	0	21		
	HE3-2	0	0	0	26	0	26		
	HE4-2	0	0	0	27	0	27		
3rd (10)	HA5-3	0	0	0	28	0	28	286 (28.6)	
	HA6-3	0	0	0	29	0	29		
	HB5-3	0	0	0	28	0	28		
	HB6-3	1	0	0	25	0	26		
	HC5-3	0	0	0	24	0	24		
	HC6-3	0	0	0	31	1	32		
	HD5-3	0	0	0	30	0	30		
	HD6-3	0	0	0	30	0	30		
	HE5-3	0	0	0	28	0	28		
	HE6-3	0	0	0	31	0	31		

APPENDIX C

Number (%) of Test Items Made and Kinds of Knowledge Measured with the Texts

(Middle School)

Grade Levels (# of tests)	Test (# of Texts Used)	Other Kinds of K.	Number (%) of Test Items Made		Total # of Test Items
			Linguistic K.	Content K.	
1st (10)	MA1-1(10)	0	7(43.75)	9(56.25)	16
	MA2-1(14)	1	5(27.78)	12(66.67)	18
	MB1-1(15)	0	7(33.33)	14(66.67)	21
	MB2-1 (7)	0	8(53.33)	7(46.67)	15
	MC1-1(12)	0	9(42.86)	12(57.14)	21
	MC2-1(11)	0	8(42.11)	11(57.89)	19
	MD1-1 (3)	0	8(100.0)	0(0.0)	8
	MD2-1 (9)	0	19(100.0)	0(0.0)	19
	ME1-1 (8)	0	10(62.50)	6(37.50)	16
ME2-1(14)	0	4(26.67)	11(73.33)	15	
2nd (10)	MA3-2(11)	0	7(38.89)	11(61.11)	18
	MA4-2 (7)	0	9(52.94)	8(47.06)	17
	MB3-2(10)	0	19(76.0)	6(24.0)	25
	MB4-2 (8)	0	9(50.0)	9(50.0)	18
	MC3-2(10)	0	15(79.95)	4(21.05)	19
	MC4-2(10)	0	6(40.0)	9(60.0)	15
	MD3-2 (9)	0	7(36.84)	12(63.16)	19
	MD4-2(13)	0	26(100.0)	0(0.0)	26
	ME3-2 (9)	0	3(23.08)	10(76.92)	13
ME4-2 (8)	0	6(37.50)	10(62.50)	16	
3rd (10)	MA5-3(10)	0	7(46.67)	8(53.33)	15
	MA6-3(16)	0	11(39.29)	17(60.71)	28
	MB5-3(14)	0	6(31.58)	13(68.42)	19
	MB6-3(11)	0	7(38.89)	11(61.11)	18
	MC5-3(15)	0	11(47.83)	12(52.17)	23
	MC6-3(14)	0	7(29.17)	17(70.83)	24
	MD5-3 (9)	0	5(29.41)	12(70.59)	17
	MD6-3(17)	0	9(37.50)	15(62.50)	24
	ME5-3(10)	0	12(63.16)	7(36.84)	19
ME6-3(13)	0	11(50.0)	11(50.0)	22	

APPENDIX D

Number (%) of Test Items Made and Kinds of Knowledge Measured with the Texts

(High School)

Grade Level (# of Tests)	Test (# of Texts Used)	Other Kinds of K.	Number (%) of Test Items Made		Total # of Test Items
			Linguistic K.	Content K.	
1st (10)	HA1-1(15)	0	14 (51.85)	13 (48.15)	27
	HA2-1(29)	0	13 (40.63)	19 (59.38)	32
	HB1-1(21)	0	10 (40.0)	15 (60.0)	25
	HB2-1 (23)	0	10 (35.71)	18 (64.29)	28
	HC1-1(24)	0	11 (42.31)	15 (57.69)	26
	HC2-1(20)	0	10 (43.48)	13 (56.52)	23
	HD1-1 (19)	0	6 (26.09)	17 (73.91)	23
	HD2-1 (26)	0	15 (57.69)	11 (42.31)	26
	HE1-1 (18)	0	9 (40.91)	13 (59.09)	22
HE2-1(26)	0	15 (53.57)	13 (46.43)	28	
2nd (10)	HA3-2(24)	0	8 (30.77)	18 (69.23)	26
	HA4-2 (29)	0	11 (37.93)	18 (62.07)	29
	HB3-2(26)	0	12 (42.86)	16 (57.14)	28
	HB4-2 (20)	0	5 (20.0)	20 (80.0)	25
	HC3-2(25)	0	9 (28.13)	23 (71.88)	32
	HC4-2(25)	0	14 (48.28)	15 (51.72)	29
	HD3-2 (21)	0	12 (46.15)	14 (53.85)	26
	HD4-2(21)	0	12 (46.15)	14 (53.85)	26
	HE3-2 (26)	0	14 (46.67)	16 (53.33)	30
HE4-2 (27)	0	13 (41.94)	18 (58.06)	31	
3rd (10)	HA5-3(28)	0	17 (51.52)	16 (48.48)	33
	HA6-3(29)	0	13 (43.33)	17 (56.67)	30
	HB5-3(28)	0	12 (42.86)	16 (57.14)	28
	HB6-3(26)	0	24 (77.42)	7 (22.58)	31
	HC5-3(24)	0	9 (31.03)	20 (68.97)	29
	HC6-3(32)	0	16 (48.49)	17 (51.52)	33
	HD5-3 (30)	0 0	11 (36.67)	19 (63.33)	30
	HD6-3(30)	0	18 (60.0)	12 (40.0)	30
	HE5-3(28)	0	15 (50.0)	15 (50.0)	30
HE6-3(31)	0	13 (40.63)	19 (59.38)	32	

Sung-Ae Kim
Dept. of English Language Education
College of Education, Pusan National University
Jangjeon-dong, Keumjung-gu, Busan, 609-735, Korea
Tel: (051) 510-2611
Email: sngkm@pusan.ac.kr

Received 15 October 2011
Revised 30 November 2011
Accepted 13 December 2011