



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

Effects of Students' Feedback on English
Teachers' Test Preparation and Administration

학생들의 피드백이 영어 교사들의 시험 준비와
시행에 미치는 영향

2012년 8월

서울대학교 대학원

외국어교육과 영어전공

정 현 진

Effects of Students' Feedback on English Teachers' Test Preparation and Administration

by
Hyun-Jin Jeong

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the
Graduate School of Seoul National University

August 2012

Effects of Students' Feedback on English Teachers' Test Preparation and Administration

학생들의 피드백이 영어 교사들의 시험 준비와
시행에 미치는 영향

지도교수 권 오 량

이 논문을 교육학 석사 학위논문으로 제출함

2012년 8월

서울대학교 대학원
외국어교육과 영어전공
정 현 진

정 현 진의 석사학위논문을 인준함

2012년 8월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Effects of Students' Feedback on English Teachers' Test Preparation and Administration

APPROVED BY THESIS COMMITTEE:

BYUNGMIN LEE, COMMITTEE CHAIR

TAE-YOUNG KIM

ORYANG KWON

ABSTRACT

The present study investigated the quality of teacher-made tests in one specific school and explored how it could be improved with the students' feedback on the test. Specifically, the study attempted to examine 1) how students react to the quality of teacher-made tests, 2) how the students' feedback influences their teachers' awareness and practice on testing, and 3) how the teacher-made tests improve after the students' feedback. About 760 students in the tenth grade took midterm and final exams, and among them, 115 students answered the questionnaires on both tests. Four Korean high school English teachers produced the exams and they had two feedback sessions in which they discussed the test results and students' feedback. The primary purpose of this study was to investigate whether the teachers' testing awareness levels and practices had changed, affected by the students' feedback.

The overall findings from the study are as follows:

1. Students' attitudes to the test fairness in both tests were not positive despite the significant improvement in the final exam. Perceived item difficulty, validity, form appropriacy and written comments by students gave useful insights for teachers to revise their tests. Compared to the multiple-choice items, students commented more on the supply-type items, which indicates that the supply-type items needed more revision and improvement.

2. Teachers' awareness and practices on testing showed significant changes throughout the study. Due to the students' feedback on the midterm exam, they

found the weak points of their test and discussed how to change them. By practicing what they discovered on the midterm exam, they saw some improvement in the quality of the final exam. Teachers' attitudes towards the feedback sessions were very positive in that they realized how their tests were perceived, understood how to change them, and recognized the results that can be brought forth by their efforts.

3. There were some improvements in the quality of the multiple-choice items in the item analyses of the final exam. This indicated that the students' feedback indeed helped the teachers improve their test. On the other hand, teachers' efforts to raise students' scores on the supply-type items did not bring forth the expected results. Overall, many items constructed by the teachers were too easy or too difficult, indicating the necessity for more revision and modification by the teachers.

These findings suggest that student feedback can be a useful tool for teachers to improve their tests. Thus, there should be more efforts among English teachers to obtain students' feedback on exams and utilize it. The need for teachers to recognize the importance of testing and continuous and tailored teacher training on assessments is also suggested.

Key Words: students' feedback, improving tests, teacher-made tests, item analysis, teachers' awareness and practice on testing

Student Number: 2008-21572

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS.....	iii
LIST OF TABLES	vii
CHAPTER 1. INTRODUCTION	1
1.1. The Purpose of the Study.....	1
1.2. The Significance of the Study.....	3
1.3. Research Questions	4
1.4. Organization of the Thesis	5
CHAPTER 2. LITERATURE REVIEW	6
2.1. Theoretical Background.....	6
2.2. Research on Teacher-made Test	10
2.3. Research on Test-taker Feedback	14
CHAPTER 3. METHODOLOGY	18
3.1. Context of the School.....	18
3.2. Participants.....	21
3.2.1. Test Writers	21
3.2.2. Test Takers	22

3.3. Materials	23
3.3.1. The School Exam Papers	23
3.3.2. The Logs of the Feedback Sessions by the Teachers	25
3.3.3. Questionnaire for the Students.....	25
3.3.4. Questionnaire for the Teachers	27
3.4. Procedures	28
3.4.1. The Midterm Exam Preparation and Administration	28
3.4.2. Survey of the Students' Feedback on the Midterm Exam	29
3.4.3. The First Feedback Session by the Teachers	29
3.4.4. The Final Exam Preparation and Administration	30
3.4.5. Survey of the Students' Feedback on the Final Exam.....	30
3.4.6. The Second Feedback Session by the Teachers.....	31
3.5. Data Analysis	31
CHAPTER 4. RESULTS AND DISCUSSION.....	33
4.1. The Teachers' Initial Stages of Testing Beliefs and Practices	33
4.1.1. The Teachers' Evaluation of Themselves as Test Writers.....	34
4.1.2. Problems and Concerns about Test Writing.....	35
4.1.3. Experience in Teacher Training on Test Development	36

4.2. The Midterm Exam	37
4.2.1. The Teachers' Views on the Quality of the Midterm Exam.....	37
4.2.2. The Students' Feedback on the Midterm Exam	38
4.2.3. The First Feedback Session	46
4.3. The Final Exam	49
4.3.1. The Teachers' Views on the Quality of the Final Exam	50
4.3.2. The Students' Feedback on the Final Exam.....	51
4.3.3. The Second Feedback Session	59
4.3.4. The Teachers' Views on the Effects of the Feedback Sessions	62
4.4. Results of the Statistical Analyses of the Exams	63
4.4.1. Descriptive Statistics of the Exam Scores.....	64
4.4.2. The Difficulty Indices of the Midterm and Final Exams	65
4.4.3. The Discrimination Indices of the Midterm and Final Exams	67
4.4.4. Distracter Analysis of the Midterm and Final Exams	70
4.4.5. Summary of the Findings	74
CHAPTER 5. CONCLUSION	76
5.1. Summary of the Research Findings and Implications.....	76
5.2. Limitations of the Study and Suggestions for Future Research.....	81

REFERENCES	83
APPENDICES	88
ABSTRACT IN KOREAN	112

LIST OF TABLES

Table 3.1 Profiles of the Teacher Participants.....	22
Table 4.1 Descriptive Statistics of Test Fairness on the Midterm Exam.....	39
Table 4.2 Means of the Perceived Item Difficulty, Validity, and Form Appropriacy on the Midterm Exam.....	40
Table 4.3 Descriptive Statistics of Test Fairness on the Final Exam.....	51
Table 4.4 Two-Way Repeated Measures ANOVA of Test Fairness.....	52
Table 4.5 Means of the Perceived Item Difficulty, Validity, and Form Appropriacy on the Final Exam.....	53
Table 4.6 Descriptive Statistics of the Midterm and Final Exams.....	64
Table 4.7 Item Difficulty Indices (p) of the Midterm and Final Exams.....	65
Table 4.8 Item Evaluation by the Item Difficulty Index for the Two Tests.....	66
Table 4.9 Item Discrimination Indices (D) of the Midterm and Final Exams.....	68
Table 4.10 Item Evaluation by the Item Discrimination Index for the Two Tests.....	69
Table 4.11 Percentage of Selecting Each Option and Degree of Distracter Attractiveness of the Midterm Exam.....	71
Table 4.12 Percentage of Selecting Each Option and Degree of Distracter Attractiveness of the Final Exam.....	72

CHAPTER 1.

INTRODUCTION

The present study mainly focuses on effects of students' feedback on English teachers' test preparation and administration in a formal testing situation in Korea. Rather than just examining the quality of the tests constructed by English teachers, the study attempts to investigate how the teachers are affected by their students' feedback on the tests and how the quality of test improves after it.

This chapter introduces the present study by providing its purpose and significance. Section 1.1 states the purpose of the study and section 1.2, the significance of the study. Section 1.3 poses the research questions. Finally, section 1.4 illustrates the organization of the thesis.

1.1. The Purpose of the Study

While a teacher's principal job is most certainly teaching their students, assessing students in an appropriate manner has become a task of the utmost importance. In the Korean educational context, university admission systems have gone through many changes, and universities are now selecting their students in various ways besides the scores of the Korean College Scholastic Ability Test (CSAT). Among these ways, the scores of tests that students took in high school have become one of the important criteria. However, the tests produced by school teachers have received much less attention in the Korean

educational system than the CSAT.

Backwash, the effect of testing on teaching and learning, is so powerful in the teaching and learning context of Korea that the teacher should ensure that the tests measure what they are supposed to measure (Kim, 2009). Considering that midterm and final exams in secondary schools are achievement tests that examine students' achievement in a subject, these tests should test the most important knowledge and skills that are taught in the class. At the same time, they should be able to rank students according to their scores, which play an important role when applying for a university. These two roles of tests are somewhat contradictory and make it difficult for teachers to always write good items. The fact that the students are taught by different teachers but take the same test in most schools also prevents teachers from assessing what they emphasized in the class.

In the Korean educational context, with numerous students and where teachers are often caught in a time crunch before midterms and finals as they attempt to create exam papers while dealing with other challenging tasks, it is never easy to produce a test of good quality. Moreover, as testing has not been a core subject in pre-service or in-service teacher training programs, teachers do not have much expertise in testing. For these reasons, teacher-made tests cannot be fully trusted in terms of their quality, and several studies have shown that tests constructed by Korean English teachers need some improvement and revision (e.g., Im, 1996; Kim, 2009).

Collecting feedback from test takers can be a powerful means of

improvement. Test takers experience the test in the most meaningful way and thus, their opinions and attitudes towards the test can provide valuable insight to those who are concerned with test development and improvement. However, efforts to collect test takers' feedback have not been very active in most testing situations, despite its importance (e.g., Bachman & Palmer, 1996; Popham, 1999).

The purpose of this study is to investigate the quality of teacher-made tests in one specific school and to explore how it could be improved by student feedback on the tests. In addition to quantitative analysis of the test results, this study also conducts a qualitative analysis of the teachers' changes in their awareness levels and practices as regards testing.

1.2. The Significance of the Study

In terms of quality, teacher-produced tests have long been a topic of concern among educators. However, despite the fact that the characteristics of teacher-made tests have been well presented in many studies (e.g., Gronlund, 1985; Popham, 1990; Soranastaporn et al., 2005), relatively few studies have investigated how to improve the quality of teacher-produced tests (e.g., Coniam, 2009; Johnson et al., 1999). In those studies which examined the quality of teacher-produced tests, the focus was on the classical test statistics and item analyses rather than feedback from the 'real-life' test takers.

In the Korean EFL situation, only a few studies have assessed how teachers

actually construct a test. There has been little, if any, empirical research that attempted to examine the effect of test-takers feedback on improving the quality of tests by English teachers in Korea.

This study will do a thorough analysis of test-taker feedback and test statistics to determine whether improvements were made to the quality of the tests and to determine the characteristics of the improvements if they exist.

The results of this study will be beneficial in at least three ways. First, through the attempt to receive students' feedback on the test, this study will enlighten English teachers regarding the importance of obtaining students' feedback to improve their test quality, against the common practice that does not consider any backwash for the future test in most schools. Second, this study will provide a means of gathering students' feedback and analyzing test items in a systematic way. Last, the study will bring about a discussion concerning the need for teachers and those concerned with education to recognize the importance of language testing and to refine assessment training programs for pre- and in-service teachers.

1.3. Research Questions

In order to investigate effects of students' feedback on English teachers' test preparation and administration, the following research questions are posed.

1. How do students react to the quality of teacher-made tests?

2. How does the students' feedback influence their teachers' awareness and practice on testing?

3. How do the teacher-made tests improve after the students' feedback?

1.4. Organization of the Thesis

This thesis is comprised of five chapters. Chapter 1 introduces the purpose and the significance of the study, and addresses the research questions. Chapter 2 reviews the theoretical background for improving tests and previous research on teacher-made test and test-takers' feedback. Chapter 3 describes methodology employed in the present study, and Chapter 4 reports the results and discusses the findings with regard to the research questions addressed. Finally, Chapter 5 concludes the thesis with a summary of major findings and their implications, as well as the limitations of the present study and suggestions for future research.

CHAPTER 2.

LITERATURE REVIEW

This chapter reviews the theoretical frameworks and relevant studies on which this study is based. This chapter begins with the discussions about the importance of tests, qualities of tests, and ways of improving tests. Section 2.2 reviews studies on teacher-made tests and section 2.3 presents studies on test-taker feedback.

2.1. Theoretical Background: Tests, Qualities of Tests, Improving Tests

Tests, which play a major role in education, can help make significant decisions about students and affect students' motivation to pursue their learning. Their impact goes beyond the measurement they attempt to fulfill (Shohamy, 1982). Such an important role of tests has been presented in a lot of research. For example, Madsen (1983) mentioned that well-constructed tests can give students a sense of accomplishment and a feeling that the teachers' evaluation of them matches what they have taught them, thus creating positive attitudes towards instruction. Madsen also stated that "good English tests help students learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve." Thus, it is necessary for teachers as well as students to try to get most out of them (Hutchinson & Waters, 1987).

There have been many attempts to evaluate and improve tests. One way of assessing tests is to examine the qualities that determine their effectiveness. For this, many researchers have provided the qualities of tests on their own. Bachman and Palmer (1996) defined these ‘good qualities’ as reliability, validity, authenticity, interactiveness, wash-back impact, and practicality. Popham (1990) presented seven evaluative factors: description of measured behavior, items per measured behavior, scope of measurement, reliability, validity, comparative data, and absence of bias. Brown (2004) also suggested five principles of practicality, reliability, validity, authenticity, and washback as useful guidelines for evaluating an assessment. It should be noted that although these researchers used different terms, the qualities of tests presented above were similar to each other. There is considerable empirical research on these characteristics – especially on reliability and validity – reflecting efforts to evaluate and improve existing tests.

While previous research guided how to evaluate a test by examining various aspects of it, research on improving test items specifically has also been carried out.

Popham (1990; 2005) defined *empirical* and *judgmental* techniques for improving test items. The former indicates relying on the item writers or examinees reviewing items according to certain standards. Students can be asked to comment if they find items ambiguous, misleading, too hard, too easy, and so on. He argued that the data gathered could be very beneficial for prospective tests. On the other hand, the latter uses examinee-response data, that is, the examinees’ actual responses to the test items. Item analyses that include item difficulty

indices¹, item discrimination indices², and distracter analysis³ can be very useful to examine an item's quality. He recommended employing both strategies to maximize the effect of improvement.

Similarly, Reynolds, Livingston, and Willison (2006) suggested using qualitative item analysis as well as quantitative item analysis to improve items. They argued that although the qualitative procedures have not received as much attention as the quantitative ones, it is often beneficial to use both of them. The argument for qualitative item analysis was mostly based on Popham's (2000) study, recommending 1) test developers' reviewing the test after a few days of constructing it, 2) having a colleague review the test, and 3) having the examinees provide feedback on the test.

Worthen (1999) also asserted the importance of discussing test results with students. He argued that such a discussion not only gives them an opportunity to "let off steam," but often results in substantially improved test items and a

¹ An item difficulty index is the percentage of students who respond correctly to it. Often referred to these days simply as a *p* value, it is calculated as follows:

$$\text{➤ Difficulty } p = R/T$$

R = the number of examinees responding correctly to an item

T = the total number of examinees responding to the item (Popham, 1999, p. 273)

² An item discrimination index tells us how frequently an item is answered correctly by those who perform well on the total test. A positively discriminating item indicates that an item is answered correctly more often by those who score well on the total test than by those who score poorly on the total test (Popham, 1999, p. 275).

³ In the case of multiple-choice items, we can gain further insights by carrying out a distracter analysis in which we see how the high and low groups are responding to the items' distracters (Popham, 1999, p. 278).

valuable learning experience for them (p. 263). Although students' feedback about the test can be useful, it should be noted that the items they dislike the most are not necessarily bad. Therefore, he recommended that student feedback should be considered in conjunction with the results of test item analysis.

Alderson, Clapham, and Wall (1995) presented a term '*post-test reports*' in which institutions keep records of their decisions, procedures, the analyses that they conduct on test results and the feedback they receive. The most important data to gather are the item results including descriptive statistics for the whole test and each of its components, and item analysis for each item. Also, they argued that feedback should be collected from administrators, examinees, and examiners regularly, by using questionnaires which ask about specific features of the test. They asserted that the post-test report can be an effective way of reminding teachers of what they need to do to prevent problems in the future tests.

All of the studies presented above shared similar procedures and methods of improving test items. They all valued feedback by test producers and test takers along with test item analysis. Recognizing that such data provide evidence of test quality, those responsible for testing should make every effort to gather information on the existing tests and make suitable adjustments to items and other facets of a test (Alderson et al., 1995).

2.2. Research on Teacher-made Test

Teacher-created tests have long been a topic of concern in terms of quality in the field of educational research. For example, Gronlund (1985) indicated that the quality of test items on standardized tests was generally higher than that on teacher-made tests, as items were written by specialists, pretested and refined, thus consequently found to be very reliable. Soranastaporn, Chantarasorn, Suwattananand, Janvitayanujit, and Suwanwongse (2005) supported the argument further, discussing a study in Thailand, which compared the concurrent validity of achievement tests produced by Thai English language teachers against commercial standardized tests. The study demonstrated low correlations between the teacher-produced tests and standardized tests, providing three reasons: first, the tests were too difficult or too easy; second, the tests measured content that had not been taught in class; and third, the tests did not show what students had actually achieved.

Regarding these concerns about tests constructed by teachers, Popham (1990) pointed out lack of time and resources as major reasons. He argued that teachers are unlikely to be skilled in test construction techniques, mainly because they are not completely aware of the principles of educational measurement (2001, p. 26). Similarly, Carter (1984) indicated in his study that teachers spent little time editing or revising tests. An analysis of interviews with 310 teachers revealed a high level of insecurity among the teachers regarding their knowledge of good item writing principles and practices commonly used in item writing (p.

59). He concluded his research suggesting that the present pre-service testing courses should be monitored whether the content coincides with testing activities at the local district and classroom level, and that teacher in-service programs should be focused on enhancing teachers' knowledge of testing and on enabling them to discuss different testing issues.

There have been some efforts to improve teacher-made tests in the form of assessment training programs. Kirschner, Spector-Cohen, and Wexler (1996) described a workshop designed to raise the teachers' awareness of potential obstacles to student comprehension of test questions. Fifty EFL teachers were assigned to groups of five or six, produced a ten-question reading comprehension test in each group, received comments from other groups, and finally revised their original test based on the comments and checklist. The study indicated the need for a test-writing workshop, addressing many advantages such as a nonthreatening atmosphere, a process of discovery, and confidence in the group process.

In Johnson, Becker and Olive's (1999) study, teachers revised their tests using different methods. The study described a Master's level language testing program for ESL teachers, where teachers-in-training produced an objective reading comprehension test which was tried on 14 subjects enrolled on a local ESL program. These teachers analyzed the tests themselves according to the classical test theory and refined their tests on the basis of their own students' responses. The study emphasized the need for a language testing course to be incorporated into local teaching and testing situations, and consequently to be

significant for participants.

Similarly, Coniam (2009) examined the quality of tests that Hong Kong primary and secondary teachers produced for their EFL students. The study examined the effects of a language testing program on graduate student teachers, where participants produced objective tests, proceeding through the stages of test specification, moderation, item analysis, and test refinement. Comments from the participants showed a considerable amount of awareness of test principles and insight into the process of test development and analysis. The study emphasized the need for the empowerment of teachers through support and training in the principles of assessment.

In the Korean EFL situation, quite a few studies have been conducted on testing for the last two decades but they were mainly on performance tests due to the increasing concern on the communicative teaching and testing during the period. There were not many studies on how school teachers actually construct test items (Kim, 2009).

There are a few studies that attempted to evaluate school achievement tests (e.g., Chae, 2004; Hong, 2009; Jung, 2004; Kim, C., 2003; Kim, S., 2005; Ko, 2007; Park, 2010). The methodologies employed in these studies varied from indicating errors, performing item analysis, comparing with standardized tests, to investigating and categorizing test contents, or analyzing the test according to the assessment guidelines in the curriculum.

The poor quality of school achievement tests was indicated in a few studies. In Im's (1996) study, ten randomly sampled teacher-made tests were analyzed

according to the criteria based on the principles of language testing. Also, the questionnaires regarding language testing practices were completed by 100 English teachers. The results indicated that teacher-made tests needed to be corrected in many ways. Similarly, Kim (2009) investigated the quality of multiple-choice test items constructed by middle school English teachers. Fifteen English tests were analyzed to determine the weaknesses of the items, examining the content, structure, and options. According to the results, the multiple-choice test items constructed by the middle school English teachers could not be fully trusted and could not contribute to achieving the beneficial backwash.

While previous two studies gathered 10 to 15 English tests from secondary schools and investigated their quality, Jung (2007) focused on one specific test performed in a high school and analyzed it more thoroughly. She used item analysis, qualitative responses from the teachers, and correlation with a standardized test as means of analyzing the test. The results indicated some improper items in the test and the need for teachers' continuous effort to improve their tests.

Recognizing the poor quality of teacher-produced tests and the need for professional training, Jeon, Park, Ahn, Oh, Yu, Lee and Kim (2005) attempted to establish the basis for professional standards of assessment by investigating Korea's current state in student assessment and by reviewing literature on student assessment and professional standards of language testers. They confirmed that both pre- and in-service teacher training programs have not provided sufficient knowledge or opportunities for student assessment. The study concluded that

efforts are necessary to develop teacher expertise in testing through well-established professional standards and supports.

2.3. Research on Test-taker Feedback

The importance of collecting feedback for test improvement has been emphasized by many researchers. According to Bachman and Palmer (1996), the primary purpose of collecting feedback is to provide information about the qualities of tests to make revisions. They emphasized the importance of collecting feedback, stating that “the more care taken to develop a test and the more feedback obtained on usefulness, the more useful it will be (p. 240).” They also argued that the process of obtaining feedback to improve test usefulness should continue as long as the test continues to be used. Popham (1990) also pointed out that when attempting to improve test items, a rich source of data is often overlooked because we typically fail to acquire advice from examinees. Yet, because examinees have experienced test items in the most meaningful context, examinee judgments can provide useful insight concerning particular items and other features, such as the test’s directions and the time allowed for completing the test (p. 272).

There has been considerable research on test-taker feedback to various types of language tests. These studies have revealed that many factors influenced student reactions to a test, including various aspects of the test, testing situation, and individual examinee differences (Bradshaw, 1990; Scott, 1986; Shohamy, 1982, 1983; Zeidner, 1988; Zeidner & Bensoussan, 1988). Examinee feedback

concerning various critical features of educational tests may be one of the most valuable sources of information about a test's subjective quality and can be highly useful to test development and revision (Shohamy, 1982; Zeidner & Bensoussan, 1988; Zeidner, 1990). Despite a great deal of research concerning test-taker attitudes and reactions, research concerning the practical application of test-taker feedback in the test construction process itself is little.

One of such studies is Alderson's (1988), which documented progress on the ELTS (English Language Testing Service) tests revision project. The study called for the gathering of feedback on draft items from test candidates to enable test developers to prepare final specifications for their tests and revise the sample items.

Kenyon and Stansfield (1991) provided a detailed methodology for gathering feedback – both quantitative and qualitative – in order to validate oral proficiency assessment tasks. Pointing out that few performance-based tests systematically collected examinees' comments during the test development stage, their study proved that paying attention to examinees' comments significantly improved the test product.

Brown (1993) investigated the use of test-taker feedback in the development of an occupational foreign language test (Japanese). She gathered the test-takers' attitudes to the test using a Likert-type rating scale and assumed that this feedback, combined with item analysis statistics, would help with selection and revision of items for the final version of the test. The findings proved the value of feedback provided by the test candidates in terms of its use

in the test revision stage and in examining the acceptability and fairness of the test from the test-takers' point of view.

In the Korean EFL context as well, only a few studies have been undertaken about the application of test taker attitudes in the test development process.

Choi (1997) used test takers' feedback as a useful tool for investigating validity of a newly developing general English proficiency test (later named as TEPS). Perceived validity by test takers was collected for each section of the test using five-point Likert-scale. He argued that the judgmental opinions of prospective test takers regarding the test method facets merit in many ways and should be incorporated in developing a fair and valid test. In his later study, Choi (1999) again investigated test fairness and validity of TEPS (Test of English Proficiency, developed by Seoul National University), based largely on test-takers' qualitative and quantitative feedback on the pilot TEPS and the first administrated TEPS. Questionnaires with five-point Likert-type scales were also used in this study. The findings showed that the majority of the test-takers considered TEPS a valid test of desirable test methods and a potential candidate to substitute for other conventional EFL tests.

There was also a research concerning test takers' perceptions regarding a spoken test. Shin, Kim, and Kang (2010) investigated Korean examinees' perceptions of testing spoken English OPIc (Oral Proficiency Interview-computer). Questionnaires with five-point Likert-type scales were used to obtain test takers' information pertaining to the test, their preparation and overall attitudes towards the test.

Another study was conducted by Joo (2008) to investigate face validity of direct and semi-direct oral tests from the test taker perspective. Compared to previous studies which used questionnaires as a main tool, she used an interview method, indicating that “the questionnaire may become too simplified or draw out even wrong or inaccurate thoughts and feelings and thus possibly miss important data.” It may have been possible due to the small size of the study’s participants (n=8). According to the interviews, test takers considered the direct oral test as a more authentic and valid test.

Unfortunately, none of these studies have explicitly examined how test takers' feedback influence teachers' test construction practices and their testing awareness. Furthermore, in the Korean EFL context, most of the studies related to teacher-made tests were on the evaluation of the tests by means of item writing guidelines or statistical analysis, excluding how the test takers responded to and felt at such tests. Therefore, the present study aims to investigate both the quality of teacher-produced tests and the effects of students' feedback on teachers' test preparation and administration.

CHAPTER 3.

METHODOLOGY

In this part of the thesis, the method of collecting the students' feedback and discussing the feedback with the test results in a high school testing situation is described. Section 3.1 introduces the context of the school to help understanding of the research. Sections 3.2 and 3.3 provide details of the participants and materials employed in the present study. Section 3.4 explains the procedures used for this study. Finally, section 3.5 describes how the data were analyzed.

3.1. Context of the School

Given that the present research was conducted on the basis of a real educational situation, knowing the context of the school would increase the level of understanding of the research overall and the results of it.

This school, located in the southern area of Gyeonggi Province, is a public high school. In this area, ninth-grade middle school students have to take an entrance exam to apply for a high school at the end of the year. High schools select their own students based on the scores on this exam and on the grades that students obtained in their middle schools. While many advanced students in this area applied for private high schools, the school that participated in the present study was a describable public high school among the students. The majority of students who applied for this school were those who ranked in the middle or

high-middle level of their class in middle school. The scores on a national achievement test that these students took after their entrance mostly ranged from grade 4 to 6⁴, which means that there were not many advanced or poor students in this school. This is a very important fact for this study because students at these levels usually depend on their grades on school exams rather than CSAT scores when they apply for a university.

Near the school, there were several private institutions. However, many students stayed in the school until late at night, studying by themselves, and only a few students attended these institutions after school. The school was located in the residential area with many apartments and restaurants. Many students came from these nearby houses, and most could be said to be from ordinary families. Parents and students associated with this school tended to depend on the school education heavily compared to wealthier areas in Korea, such as Gangnam in Seoul. Therefore, they were more cooperative with the school and paid great attention to school education.

The tenth-grade students learned English four periods a week. Three periods were taught by Korean English teachers studying the reading and grammar parts of the textbook, while a native-speaking English teacher taught the listening and speaking sections, for one period. The students learned only the textbook during regular periods, but if they wanted, they could learn more by studying a private reference book in an after-school class. The size of the school was very large in

⁴ The entire range is from grade 1 (high) to 9 (low).

that there were 16 classes in the tenth-grade, with each class composed of approximately 40 students. For effective teaching, each student was placed in high-, mid-, or low-level classes on the basis of their English score on the placement test. Therefore, the classes were rearranged during the English periods, with four high-level classes, eight mid-level classes and four low-level classes overall. Actually, this proficiency-based class brought about many issues among the teachers because it had both advantages and disadvantages. While it helped the students to learn effectively based on their own level, it could harm the atmosphere of class as the students were from different home classes and did not know each other well.

Four Korean English teachers and one native-speaking English teacher taught English to tenth-grade students. The four Korean English teachers did not know each other well at the beginning of the year, but this was easily rectified thanks to their similar features, such as their genders and ages. They discussed the teaching content, tests, or materials frequently and were passionate about their teaching. Meanwhile, the native-speaking English teacher was a new teacher in this school. It was her first year to teach English in Korea, though she had taught in the USA before. She taught the speaking and listening parts of the textbook with the cooperation of a Korean teacher. The numerous students and her newness made it difficult for her to teach speaking and listening effectively. When she taught, a Korean English teacher stayed in the classroom with her, helping her or managing the class. Although the native-speaking teacher led the class, Korean teachers wrote test items for the listening and speaking parts as

they has more experience in this area.

3.2. Participants

There were two groups of participants in the present study: test writers and test takers. Sections 3.2.1 and 3.2.2 provide details about test writers and test takers, respectively.

3.2.1. Test Writers

The test writers in the present study were four high school English teachers (including the researcher) who currently work at a high school in Gyeonggi Province, Korea. These four teachers were teaching English to tenth graders and producing school achievement tests – midterm and final examinations – each semester. These teachers had more than two years of English teaching experience, and all were female. They had no experience in evaluating their own tests by collecting students' feedback, but showed strong interest in improving the quality of their tests.

The researcher was one of these English teachers. In her school, when teachers produce tests, one teacher edits the overall format, and the others produce items. In the present study, the researcher took charge of editing for two achievement tests so as not to affect the research results. Profiles of the teacher participants are presented in Table 3.1.

Table 3.1
Profiles of the Teacher Participants

Test writers	Teaching experience (years)	Testing experience (no. of times)
Teacher A	4	16
Teacher B	2	5
Teacher C	2	4
Teacher D (= Researcher)	4	20

*Testing experience refers to how many midterm, final, or other important exams she had produced in her teaching career.

3.2.2. Test Takers

The test takers in this study were tenth grade high school students at the same school mentioned above. In the present study, there were two groups of student participants. Initially, approximately 760 tenth graders took the midterm and final exams and their scores were analyzed. Secondly, among them, 115 students responded to the questionnaire both on the midterm and the final exams. In order to obtain feedback from students of different proficiency levels, 41 high-level, 36 mid-level and 38 low-level students participated in answering questionnaires. Thus, one advanced, one mid-level, and one low-level class took part in the survey. These classes were taught by the researcher at the time. The researcher selected

her own class as respondents, as she assumed that students in these classes would be more cooperative than other students in responding to somewhat lengthy questionnaires. The students who did not answer some of the questions were again asked to complete the questionnaire by the researcher. Thus, there were no missing data in the two questionnaires. Of the 115 students, the male-female ratio was 61:54. The subjects had no experience in giving feedback on the tests they had taken as formally and thoroughly as conducted in this study.

3.3 Materials

This section presents materials for the present study. Section 3.3.1 presents the school exam papers used in this study and section 3.3.2 describes the logs of the feedback sessions by the teachers. Sections 3.3.3 and 3.3.4 introduce the questionnaire for the students and for the teachers respectively.

3.3.1. The School Exam Papers

The midterm and final exams produced by the participating teachers were analyzed in the present study. Test production and administration processes were all identical in the two tests. The tests consisted of 23 multiple-choice items (hereafter MC items) and five supply-type items⁵. The types of questions

⁵ Items can be grouped into the selective type and the supply type categories according to the examinees' response type. Supply-type questions again can be grouped into completion form (filling in a blank in a sentence or paragraph), short-answer form (answering the given question

comprising the MC items were similar to those on the CSAT, and the scores, ranging from 3.0 to 4.0, were assigned differently according to the teachers' judgments of difficulty. The supply items required students to write down a few words or a sentence in English. These were assigned four points each. The full scores for the MC items and the supply items were 80 and 20, respectively. The teachers collaborated in producing the test paper. Three chapters of the English textbook were covered for each test. Therefore, each chapter was assigned to one of the three teachers and 10 to 11 test items were collected for each chapter in each case. These items were put together and discussed by the teachers. They revised some test items and removed poor ones. When all of the teachers agreed on the final version of the exam, the researcher who did not produce the test items edited the paper for both the midterm and final exams so as not to affect the research results. The midterm and final exam papers are presented in Appendix 1.

on a short form), and essay form (expressing ones' opinion in a sentence or paragraph form). Since the year 2010, it has been mandatory to include a certain percent of supply-type questions on school exams in Gyeonggi Province. The purpose of this act is to assess the process of solving problems, and to raise students' creativity and problem-solving abilities. In reality, the completion form and short-answer form are preferred by most schools to essay-type questions because they are easier to score.

3.3.2. The Logs of the Feedback Sessions by the Teachers

To review the test results and students' responses to the questionnaire, the teachers had voluntary meetings after the midterm and the final exams. These meetings were recorded and discussed in the present study. The first feedback session was conducted in May, and the second one was in July of 2010, approximately two weeks after the administration of each test. In these feedback sessions, the teachers reviewed the test results and the students' feedback and discussed their findings. They decided on what to change to improve their own tests through these feedback sessions.

3.3.3. Questionnaire for the Students

The students' comprehensive feedback to the test was gathered using a five-point Likert-type rating scale and open-ended questions (See Appendix 2). The questionnaire, based on the previous studies on evaluating test items (Bradshaw, 1990; Brown, 1993; Popham, 2005), was designed to obtain students' responses to the midterm and final exams.

The questionnaire consisted of three sections. The first section collected basic information such as their names, gender, and class. It also contained an item about 'test fairness' to assess whether they felt the test accurately measured their English ability.

The second section asked about perceived item difficulty, validity, and form

appropriacy on each item and also allowed the students to write comments for each item. Item difficulty refers to how difficult the items were for the students. Item validity alludes to whether the item tested appropriately what was taught. This concept can also be called ‘face validity’, as it attempts to assess if the item was familiar and valid from the perspective of the students. Lastly, item form appropriacy refers to whether the form of the item was appropriate in its question type, or how it presented options or conditions.

There were a total of 28 sets of four questions in total, which meant that students had to answer the same questions recurrently 28 times. For this section, students were encouraged to take out their test paper and recall what they thought during the test. One example is shown below.

	It was difficult.	1—2—3—4—5
	It tested appropriately what was taught.	1—2—3—4—5
Item 1	The form of the question was appropriate.	1—2—3—4—5
	(Comments)	

* (1)- Strongly Disagree (2)- Disagree (3)- Neutral (4)- Agree (5)- Strongly Agree

Finally, the third section requested students’ comments and suggestions on the exam in general. The same questionnaires were used for both the midterm and final exams.

3.3.4. Questionnaire for the Teachers

In order to determine the teachers' changes in their testing awareness and practices throughout the study, the teachers were asked to complete open-ended questionnaires. The researcher used questionnaires rather than interview because she had close relationships with the participating teachers; therefore, it may not have been comfortable for these teachers to express themselves freely in a face-to-face interview situation. To offset the shortcomings of the questionnaire method, additional short interviews were carried out when necessary.

The questionnaires mainly asked the teachers to do the following:

- (1) comment on their awareness and practice on testing. (self evaluation as a test writer, problems and concerns about test writing, and experience in teacher training on test development)
- (2) consider the quality of the midterm and final exams they had just produced.
- (3) consider the feedback sessions they had experienced, and evaluate their effects on improving the quality of their own tests.

The questionnaires, adapted from Coniam's (2009) study, were administered three times throughout the study: before the administration of the midterm exam, before the administration of the final exam, and after the second feedback session. They are presented in Appendix 3.

3.4. Procedures

The procedures for the present study were carried out in six phases:

- (1) The Midterm exam preparation and administration
- (2) Survey of the students' feedback on the midterm exam
- (3) The first feedback session by the teachers
- (4) The Final exam preparation and administration
- (5) Survey of the students' feedback on the final exam
- (6) The second feedback session by the teachers

3.4.1. The Midterm Exam Preparation and Administration

The midterm exam was administered at the end of April of 2010. From early April, the four English teachers produced the midterm exam questions in their usual manner. First, each teacher produced test items for each chapter they had been assigned. Then, those items were collected and discussed by the teachers. They removed some items and revised some others. The test was continuously reviewed by the teachers until the midterm exam administration. While working on the test production, the teachers completed questionnaire I regarding their testing beliefs and practice and the items they produced for the midterm exam. The midterm exam was administered on April 28th, 2010 and a total of 763 students took the test.

3.4.2. Survey of the Students' Feedback on the Midterm Exam

One week after the midterm exam, 115 students selected from among all test takers completed the questionnaire on the midterm exam (See Appendix 2). This questionnaire was designed to garner information about the quality of the test from the students' perspective. It mainly asked the students to answer questions on the topics of test fairness, item difficulty, validity, and form appropriacy. Because the students were unfamiliar with these concepts, the researcher briefly explained them before writing the questionnaire. The students had enough time to complete the questionnaire in a comfortable and relaxing atmosphere. The announcement that their opinions and suggestions on the test may be applied to the final exam motivated the students to participate carefully.

The answers were collected and analyzed using SPSS 18.0. At the same time, the students' actual answers on the test were analyzed, specifically for item difficulty, item discrimination, and a distracter analysis using the same tool.

3.4.3. The First Feedback Session by the Teachers

Two weeks after the midterm exam, the teachers who had participated in producing the midterm exam had a meeting in order to discuss the results of both the test and the students' feedback on the test. Through the meeting, the teachers were expected to become aware of how well they constructed the test items, what problems they had, and what they had to do in order to improve the test quality.

The researcher had prepared the necessary data in advance, and the meeting lasted for one hour.

3.4.4. The Final Exam Preparation and Administration

In the middle of June of 2010, the teachers started producing the final exam paper using their newly gained knowledge from the first feedback session. The procedures for test production were similar to those of the midterm exam, but the teachers exerted more effort in applying what they had learned in the feedback session by discussing the test items more frequently and thoroughly. Questionnaire II was completed to investigate the teachers' test preparation during this period. The final exam was given on July 6th, and 767 students took the test.

3.4.5. Survey of the Students' Feedback on the Final Exam

One week after the final exam, the students who participated in providing feedback on the midterm exam completed the questionnaire about the final exam. The results were analyzed using SPSS 18.0. Statistical analyses of the test results were also conducted.

3.4.6. The Second Feedback Session by the Teachers

The teachers gathered to discuss the test results of the final exam and the difference in quality between the midterm and final exams. They discovered what they had changed and what they still had to change through this meeting. The teachers then completed questionnaire III to document any changes in their testing beliefs and practices as well as how the students' feedback helped them to make these changes.

3.5. Data Analysis

In recent years, language testing researchers have become more interested in and familiar with 'qualitative' research techniques, which can be used for investigating the validity of a test. One example of such a technique is the use of introspective reports from test takers and examiners of test performance (Alderson et al., 1995). Such qualitative data can supplement the results of an analysis of quantitative data by providing insight into what students and examiners were thinking. Thus, the present study employed both quantitative and qualitative data collection and analyses methods.

Statistical analysis was conducted focusing on the following concerns. All statistical analyses were run using SPSS 18.0 with the alpha level set to .05. First, for the student questionnaires which used five-point Likert-type scales, descriptive and frequency analysis including means and standard deviations were used to

summarize the students' background information and attitudes towards the midterm and final exams. In order to check if the students' reactions towards test fairness between the two exams differed according to their proficiency level, a two-way repeated measure ANOVA was performed. Second, the descriptive statistics of the test results were analyzed for both tests. For the MC test items, the item difficulty index, item discrimination index, and item response distribution for the distracter analysis were calculated through SPSS.⁶

Along with the statistical analyses, students' comments on the tests they had taken and teachers' remarks obtained by the questionnaires and through the feedback sessions were collected and reviewed. These analyses involved a qualitative method to capture students' subjective evaluations toward the tests and teachers' changes in their testing practice and beliefs.

⁶ The NEIS (National Education Information System) provides teachers with the basic statistics of a test, including the percentage of correct response and the item response distribution. However, in the present study, in order to obtain more precise and detailed information, the students' answers were run using SPSS. The students' answers were easily obtained and computed because they were all written on OMR (Optical Mark Reader) cards for school achievement tests.

CHAPTER 4.

RESULTS AND DISCUSSION

This chapter provides findings on the effects of students' feedback on English teachers' test preparation and administration based on the data collected and used in this study. Sections 4.1 to 4.3 are composed according to a time sequence to show the changes in the teachers' perceptions and practices on testing and the students' feedback as they experienced the midterm and final exams. Section 4.1 discusses the teachers' initial stages of testing beliefs and practices at the beginning of the research. Sections 4.2 and 4.3 address the midterm and final exams respectively, discussing the students' feedback and the teachers' discussions about the results. Lastly, section 4.4 reports the test statistics results on the midterm and final exams in order to determine if any improvement occurred in the quality of the final exam.

4.1. The Teachers' Initial Stages of Testing Beliefs and Practices

On questionnaire I, which was filled out before the administration of the midterm exam, the teachers were asked to comment on their awareness and practices of testing. The following sub-sections report the teachers' evaluations of themselves as test writers, their problems and concerns related to test writing, and their teacher training experience on test development.

4.1.1. The Teachers' Evaluation of Themselves as Test Writers

When asked if they were satisfied with their test writing ability, all of the teachers answered that they were capable of test writing and that the quality of their test items had continued to improve as they experienced more and more testing. However, their levels of confidence varied depending on their levels of experience. While an experienced teacher revealed her confidence in test writing in that she made less ambiguous items and more meaningful items as she grew more experienced, two teachers with less experience showed less confidence. Their comments are presented below⁷:

My testing ability keeps improving, and I can give myself 80 out of 100 points. I try to read the texts thoroughly before writing items and review continuously so that I do not make ambiguous items. (Teacher A)

Compared to the very first test that I produced, my testing ability has improved. I try to write items which require higher-level thinking ability than assessing simple knowledge or skills. I can give myself 70 out of 100 points for my testing ability. (Teacher B)

⁷ Translation is the researcher's.

I can give myself 70 out of 100 points as a test writer. When I first participated in test writing at this school, I had no idea what I should do. I now know all of the procedures of test writing and I produce better items. (Teacher C)

4.1.2. Problems and Concerns about Test Writing

The problems and concerns about test writing were various among the teachers. One teacher mentioned the problem of a proficiency-based class:

In proficiency-based classes, where the students are placed in high-, mid-, and low-level class according to their scores, it is hard to determine the proper difficulty level of the items because the students learn at different levels but take the same test. (Teacher A)

Teacher B pointed out the basic issue of test validity:

It is not always easy to assess what is supposed to be assessed. It takes time to control other variables so as not to influence item validity. (Teacher B)

The difficulty when deciding on item type and vocabulary was also mentioned:

Because three or four teachers individually write the test items on the

assigned chapter, some question types could be repeated too many times. When this happens, we have to write new items to balance the different item types. It is very time consuming. Also, when I produce options, selecting appropriate words is very difficult. If the test contains difficult words that were not learned in class, it could be unfair to students who solely depend on school education. In that case, advanced students have an advantage. (Teacher C)

4.1.3. Experience in Teacher Training on Test Development

Every teacher had different experiences in participating in classes or workshops on assessment, but they all remarked they did not learn any practical and specific knowledge or skills on testing. The classes at universities and workshops by local education commissions cannot provide tailored skills and knowledge for every teacher.

When asked whether they had ever gathered feedback from the students after an exam, everyone answered in the negative. They may have asked one or two students in a very informal way, and then looked at the percentages of the correct responses or item response distributions provided by NEIS (National Education Information System); however, no specific and detailed data had been gathered before.

The teachers participating in this study showed curiosity and enthusiasm about gathering students' feedback on exams. They anticipated that if they tried this, they could make a better test in the future.

Finally, these teachers were fully aware of the importance of testing in their career. They stated that testing is directly connected to teaching, making it as important as teaching. They wanted to build their ability in the testing field.

4.2. The Midterm Exam

This section presents the perceptions of the teachers and the students on the midterm exam and the teachers' discussions of the students' feedback and the test results. Section 4.2.1 addresses the teachers' views on the quality of the midterm exam before its administration. Section 4.2.2 reports the students' feedback on the midterm exam. Lastly, section 4.2.3 describes the discussions that took place in the first feedback session.

4.2.1. The Teachers' Views on the Quality of the Midterm Exam

Regarding the midterm exam that they had just written, all of the teachers revealed their confidence in the quality of the final draft. One teacher reported that MC items 18 and 23 may not be good items because they only tested the simple memorization of certain grammatical aspects. Regarding MC item 16, one teacher mentioned that it may be too difficult, but another teacher considered it to be a good item. As for the supply items, the teachers were not completely sure about the quality of the items. When asked to comment on what was difficult when writing the test, the teachers gave various answers:

It was the very first exam for the whole year. Thus, I didn't know the students' levels. It was difficult to decide on the difficulty levels of the questions.
(Teacher A)

It was very difficult to write good items for the assigned chapter. The text in the Lesson I was too easy, but I had to write difficult questions to discriminate the students.... I wanted to assess what I emphasized in class, but sometimes I couldn't. I had to avoid the same or similar questions that were on the test last year. (The same textbook has been used for three years.) Therefore, I had no choice but to write a question assessing less important content. I didn't like it.
(Teacher B)

Since we only used the textbook when producing the test items, it was very difficult to write good items. Usually texts in the textbook are easier and shorter compared to other texts in the private reference books or CSAT questions... When writing supply questions, it was difficult to decide on the form of the questions. It could be too easy or too difficult. I thought I needed more information about how to write a good supply item. (Teacher C)

4.2.2. The Students' Feedback on the Midterm Exam

A total of 115 answers of students (from 41 high-level, 36 mid-level, 38 low-

level students) were analyzed for both the midterm and final exams. The first section of the survey requested basic information pertaining to the respondents, asking if the test succeeded in assessing their English ability fairly enough. It was composed of five-point Likert-type scales ranging from (1), Strongly Disagree to (5), Strongly Agree. Table 4.1 summarizes the descriptive statistics of test fairness by students on the midterm exam.

Table 4.1
Descriptive Statistics of Test Fairness on the Midterm Exam

Group	Midterm exam	
	Mean	St. deviation
High (n=41)	2.44	.950
Mid (n=36)	2.89	.820
Low (n=38)	2.50	.726
Total (n=115)	2.60	.856

According to Table 4.1, the extent of test fairness was the highest in the mid-level proficiency group, followed by the low and high groups on the midterm exam. The means of test fairness were under 3.0 for all groups indicating that the students were somewhat negative about test fairness. This means that many students considered the test as not successful in assessing their English ability and that they were dissatisfied with the results.

In the second part of the survey, the students were asked to take out their test papers and answer the questionnaires. They had to answer three elements of each item: item difficulty, item validity, and item form appropriacy, on five-point bipolar scales ranging from (1), Strongly Disagree to (5), Strongly Agree. Also, a blank was given to let the students comment on each item. Table 4.2 summarizes the results for the means of the perceived item difficulty, validity, and form appropriacy of each item on the midterm exam.

Table 4.2
Means of the Perceived Item Difficulty, Validity,
and Form Appropriacy on the Midterm Exam

Item No.	Difficulty	Validity	Form Appropriacy
1	2.60	3.39	3.41
2	2.97	3.02	3.12
3	2.52	3.48	3.44
4	3.30	3.19	3.19
5	2.49	3.58	3.55
6	3.28	3.25	3.26
7	3.13	3.14	3.27
8	3.80	2.75	2.97
9	2.47	3.52	3.51
10	2.39	3.47	3.50
11	2.38	3.63	3.58

12	2.87	3.49	3.47
13	2.82	3.44	3.45
14	2.94	3.46	3.42
15	3.20	3.48	3.40
16	4.34	2.48	2.58
17	3.14	3.23	3.11
18	2.75	3.41	3.40
19	3.81	3.02	3.06
20	2.83	3.39	3.36
21	3.71	3.22	3.17
22	2.68	3.33	3.32
23	3.10	3.39	3.42
S1	3.26	3.28	3.21
S2	3.75	3.09	3.08
S3	3.44	3.30	3.28
S4	3.63	3.18	3.11
S5	2.71	3.53	3.47
Mean	3.08	3.29	3.29

*S= Supply Item

As shown in Table 4.2, the means of the perceived item difficulty, validity, and form appropriacy were 3.08, 3.29, and 3.29 respectively. Perceived item difficulty ranged from 2.38 to 4.34. The MC items 8, 16, and 19 and supply

items 2 and 4 were found to be difficult for the students. In the cases of perceived item validity and form appropriacy, most of the items had means higher than 3.0, which means that the students felt the items to be moderately valid and appropriate in their form, except for items 8 and 16.

The students responded that MC item 8 was less valid and appropriate as well as difficult. This item asked the students to match vocabulary items with their corresponding English meanings. The students may have considered this item difficult because this type of item is not included on the CSAT and because the options were all written in English. Some students commented as follows⁸:

I didn't expect this type of question would be on the exam.

I did not learn this in class.

On the other hand, MC item 16 asked the students to find appropriate words to complete a summary of a paragraph. This is one of the most difficult CSAT question types for most students. Item 16 contained ambiguous, confusing distracters, such that only a few students were able to complete it correctly. Common comments from the students were as follows:

It was too confusing.

I cannot accept the correct answer. Option D is a more appropriate answer.

⁸ Translation is the researcher's.

Too many difficult (outside of the textbook) words were included in the question.

MC item 19, which asked about topic of a paragraph, was also difficult for the students. All of the distracters were written in English and distracters A and C were very attractive. Some students commented as follows:

It contained too many difficult words.

Options were too long.

Regarding the supply items, items 2 and 4 were considered to be more difficult than others. Item 2 asked the students to select necessary words in the box and complete a sentence. Comments by the students are presented below.

I should memorize the textbook to fill in the blank because the missing part of the sentence was too long.

The form of the question was inappropriate. It was too difficult to choose necessary words in the box.

On the other hand, item 4 asked the students to compose a missing sentence corresponding to the Korean translation provided. Some students commented as follows:

The answer is too long.

There are too many different possible answers.

The teacher did not mention that this is important in class.

I should memorize the textbook to answer this question.

I need more words provided to answer this question.

In the third part of the survey, students were asked to comment freely on the test in general. The following comments were most common among the students' responses. The comments are grouped into those pertaining to the MC items, the supply items, and the test as a whole and are presented according to the students' levels to determine if the comments differed among different proficiency groups⁹.

- Comments on the MC items

- 1) Options included too many difficult words that are not in the textbook.

–M (2), L (3)

- 2) Questions asking about grammar and vocabulary are difficult. – M (2), L (2)

- Comments on the supply items

- 3) Partial credit should be given to minor errors such as spelling mistakes.

–H (3), M (3), L (4)

⁹ H, M, L denotes the high-, mid-, low-proficiency groups respectively. The numbers parenthesized at the end of each group represent the number of students who responded similarly in each group.

- 4) Without textbook memorization, it is hard to answer correctly.
– H (3), M (2), L (2)
 - 5) The types of questions should be varied besides sentence-completion.
– H (3), M (3)
 - 6) It was too difficult. – H (2), M (2), L (3)
 - 7) Some questions were out of expectation. – M (2), L (3)
 - 8) Options should be provided to help answer the question. – M (2), L (1)
- Comments on the test as a whole
- 9) Modifying some texts in the textbook is better than using the original texts as they are. – H (2)
 - 10) It was difficult. In particular, the supply items were too difficult.
– M (2), L (2)
 - 11) It was worthwhile. I feel like I got rewarded. – H (1), M (2)
 - 12) I wish the final exam would be easier. – L (3)
 - 13) Students learn in proficiency-based classes according to their English scores. Teachers and learning contents are different in these classes. Information about the exam is also given differently. It is unfair. – M (2), L (1)
 - 14) The speaking and listening sections that are taught by the native-speaking English teacher are difficult to understand. – M (1), L (2)

Many students commented on the supply items. Most of the items were sentence-completion tasks using their Korean translation presented in the question and these sentences were mostly taken from key sentences in the chapter. Essentially, in order to lessen the students' burden, students were informed in advance that supply items would come from several key sentences that they had learned. Nevertheless, only a few students could memorize the exact sentences and as a whole they did very poorly on these items. Furthermore, the scoring was too strict, as no partial credit was given. This is why many students insisted that partial credit should be given to minor errors such as spelling. Some students mentioned that the types of questions should be varied besides sentence-completion.

Students' opinions about the test as a whole varied. Some highly-proficient students remarked that the exam was easy and that they liked it when a question modified the original text in the textbook. On the other hand, mid-level students' opinions varied regarding the difficulty of the test, while some low-level students hoped that the final exam would be easier. The varying comments on a specific points show that level differences exist in the students' perceptions.

4.2.3. The First Feedback Session

Two weeks after the midterm examination, all of the English teachers gathered for their first feedback session on the midterm exam. The researcher provided the teachers with materials containing the test statistics and the students'

feedback on the midterm exam. During this feedback session, all of the information was read through and the major findings were actively discussed. A summary of the information discussed is presented below:

- 1) Test statistics: mean, standard deviation, item difficulty, item discrimination, distracter analysis
- 2) Test fairness perceived by students
- 3) Questions that students considered to be difficult or easy
- 4) Questions that appeared to be problematic, identified by inappropriate statistical indices or students' comments
- 5) Major comments on the exam by students

After examining the data derived from the midterm exam, the teachers started to add their ideas and thoughts on the findings. They concluded that they failed to obtain a proper statistical index on certain items due to a lack of consideration and concentration, a lack of communication among themselves, and different teaching content among the teachers. It was noted that too many questions measured simple knowledge rather than comprehensive inference ability. The teachers determined that when writing options, thorough consideration and cross-checking were needed to avoid ambiguity. Some content agreement and gap-filling questions were found to be easy for students because they did not have to read the given texts, which were identical to the text in the textbook. In order to control the difficulty level, the teachers agreed to alter the text slightly in the textbook by paraphrasing some

sentences.

Regarding the supply items, the teachers agreed to make some changes to raise the students' percentage of correct responses. They decided to give partial credit if only minor errors such as spelling mistakes. Actually, many students made these types of small mistakes when they copied their answers from the test paper to the answer sheet due to a lack of attention. Also, questions should be clearer and more meaningful so that students can answer the questions by exploiting their knowledge without text memorization. The teachers had problems writing supply items, as it was the first year that the education office forced school achievement tests to include a certain percentage of supply questions. Thus, these teachers needed more information about how to construct these types of questions.

Some students confessed that they had difficulty in the conversation class taught by the native-speaking English teacher. The native-speaking English teacher was responsible for the listening and speaking parts of the textbook, but many students failed to understand the learning content and answered related questions incorrectly. The conversation class needed a change to easier English and a step-by-step process so that students could achieve the goal of the class.

Another concern among the teachers was that they cannot use the words in the textbook continually when writing items. The use of some words outside of the textbook in some options could not be avoided. However, some students commented that they did not learn these words in class and that it was unfair. The teachers knew about this complaint but they considered that the words on the midterm exam were not difficult and that students should know these words at the

high school level. Still, word choice was a difficult task for the teachers because they had to predict their students' levels and thus needed assistance when they were not sure that their word choice was appropriate for their students' level.

The teachers agreed that they should discuss and review the items more actively and thoroughly to avoid ambiguous items and to measure what they taught their students. Because several teachers participated in the item writing activity, it was not always easy to assess what every teacher taught and emphasized in class.

This meeting lasted for one hour. Afterwards, the teachers agreed to practice what was discussed on the final exam.

4.3. The Final Exam

This section discusses the perceptions of the teachers and the students on the final exam and how they differed from the midterm exam. Section 4.3.1 addresses the teachers' views on the quality of the final exam before its administration. Sections 4.3.2 and 4.3.3 report the students' feedback on the final exam and the discussions generated during the second feedback session. Lastly, section 4.3.4 discusses the teachers' views on the effects of the feedback sessions.

4.3.1. The Teachers' Views on the Quality of the Final Exam

On teacher questionnaire II, collected immediately before the administration of the final exam, the teachers responded that their testing practices for the final exam had changed in various ways due to the first feedback session. The changes that teachers mentioned are presented below:

- 1) tried to spend more time in item writing
- 2) tried to review the test items more intensely and actively
- 3) tried to write attractive distracters for all items
- 4) tried to give clear directions and partial credit for supply items
- 5) tried to exclude question types that students considered too easy in order to control the difficulty of the test overall
- 6) tried to write options in English rather than in Korean in order to control the difficulty of the test overall.
- 7) tried to consider the students' levels when choosing words for item writing.

All of the teachers reported that their own test writing ability improved somewhat and that the final draft of the final exam was very satisfactory. They expected that items on the final exam would be better in the item analyses compared to those of the midterm exam but that the students' attitudes regarding test fairness may not improve because more difficult items were included on the test.

4.3.2. The Students' Feedback on the Final Exam

During the first period of the English class after the administration of the final exam, the students were asked to fill out the questionnaire on the final exam. The same questionnaire used for the midterm exam was used again for this. Table 4.3 summarizes the descriptive statistics of test fairness according to the students on the final exam.

Table 4.3
Descriptive Statistics of Test Fairness on the Final Exam

Group	Final exam	
	Mean	St. deviation
High (n=41)	3.07	.848
Mid (n=36)	2.97	1.082
Low (n=38)	2.76	.768
Total (n=115)	2.94	.911

On the final exam, the high group indicated the greatest extent of test fairness, followed by the mid- and low-level proficiency groups. On the midterm exam, it was the mid-level group which showed the greatest satisfaction, followed by the low and high groups. The change in test fairness between the midterm and final exams was the greatest in the high group. It is notable that test

fairness increased on the final exam for all groups, although the final exam was more difficult than the midterm exam considering their means. On the other hand, the mean of the test fairness on the final exam was still under 3.0, indicating that the students were not satisfied with it.

In order to investigate whether there was a significant group difference in the two tests, the two-way repeated measures ANOVA was conducted for the three proficiency groups. The results are shown in Table 4.4.

Table 4.4
Two-Way Repeated Measures ANOVA of Test Fairness

Source	DF	Type III SS	Mean Square	F Value	P
Tests	1	6.126	6.126	9.555	.003
Group	2	3.328	1.664	1.864	.160
Tests*Group	2	3.072	1.536	2.395	.096

As shown in Table 4.4, the extent of test fairness on the final exam was significantly different from that on the midterm exam overall ($p=.003$). On the other hand, the extent of test fairness by different proficiency groups was not significantly different ($p=.160$). This means that there was no significant group difference regarding test fairness in the two tests. There was no interaction effect between the tests and proficiency groups ($p=.096$).

In the same way as on the midterm exam, perceived item difficulty, validity

and form appropriacy of the items on the final exam were surveyed in the second part of the questionnaire. Table 4.5 summarizes the results for the means of the perceived item difficulty, validity, and form appropriacy of the items on the final exam.

Table 4.5
Means of the Perceived Item Difficulty, Validity,
and Form Appropriacy on the Final Exam

Item No.	Difficulty	Validity	Form Appropriacy
1	2.41	3.37	3.49
2	2.51	3.50	3.46
3	3.19	3.22	3.15
4	2.94	3.29	3.33
5	3.50	2.60	2.86
6	2.28	3.50	3.39
7	2.67	3.46	3.42
8	2.84	3.37	3.36
9	2.78	3.34	3.29
10	3.35	3.24	3.19
11	2.92	3.41	3.40
12	2.83	3.46	3.34
13	2.63	3.31	3.33
14	2.98	3.42	3.40

15	2.67	3.35	3.38
16	3.17	3.36	3.29
17	3.00	3.38	3.32
18	3.13	3.32	3.28
19	2.65	3.49	3.57
20	2.91	3.40	3.43
21	2.63	3.42	3.46
22	2.85	3.43	3.38
23	2.95	3.39	3.34
S1	2.80	3.27	3.29
S2	3.17	3.25	3.13
S3	3.03	3.08	2.88
S4	3.39	3.34	3.25
S5	3.25	3.20	3.09
Mean	2.91	3.33	3.30

*S= Supply item

According to Table 4.5, the means of the perceived item difficulty, validity, and form appropriacy were 2.91, 3.33, and 3.30 respectively. Compared to those on the midterm exam, the perceived item validity and form appropriacy remained at nearly the same levels while the perceived item difficulty decreased slightly. This result is notable because students perceived that items on the final exam were easier than those on the midterm exam despite the fact that the actual

difficulty of the final exam was higher than that of the midterm exam.

On the final exam, most items were manageable for the students, as the perceived difficulty ranged from 2.41 to 3.50. Students considered MC item 5 and supply question 4 as difficult items. All of the items achieved appropriate perceived item validity except for MC item 5 and the students reported that the forms of the question were all satisfactory except for MC item 5 and supply item 3.

MC item 5 did not have good face validity or form appropriacy according to the students. This item asked about the writer's tone of voice and, in order to answer it, students had to know vocabulary items related to one's tone of voice. It is assumed that students who were not accustomed to CSAT question types found MC item 5 difficult and very new to them. Comments from students were as follows:

I didn't learn this in class.

I did not know the words in the options.

I don't think knowing the writer's tone is important. This question was not expected.

Supply item 4 was also difficult for the students; it asked the students to compose a sentence with the same meaning as that presented using a specific grammatical form (*not only~ but also*). Students commented as follows:

The answer was too long.

It was too difficult.

I did not learn this in class.

Finally, the students responded that supply item 3 was less appropriate in its form. This item required the students to write a sentence in English corresponding to a Korean sentence using the present-perfect progressive. The difference between this item and other supply items was that the answer for this question was not the exact sentence from the textbook. Actually, teachers deleted one minor adverb *–almost–* to make the question easier, but students who memorized the sentence without understanding wrote the same sentence in the textbook with the adverb and thus lost some points. Students commented as follows:

I did not expect for teachers to change the sentence in the textbook.

I was very embarrassed when I knew that I was supposed to delete ‘almost’ in my answer.

Compared to the comments on the midterm exam, the number of comments students wrote for each item on the final exam decreased remarkably. Assuming that students must have written their comments on the items that they considered problematic, these results are desirable.

In part three of the survey, the students freely commented on the test in

general. Their comments were categorized into three parts, like on the midterm exam.

- Comments on the MC items

- 1) I got frustrated when I came across difficult words that I didn't learn in class.
– M (2), L (2)
- 2) I hope to learn more synonyms, similar expressions and sentences to prepare for the exam. – M (2), L (1)

- Comments on the supply items

- 3) It should be easier to get partial credit. The scoring is still too strict.
– H (2), M (3), L (3)
- 4) I think that sentences taken for the supply questions should be more emphasized and practiced during class. – M (2), L (2)
- 5) They should be easier. –M (2), L (3)
- 6) The questions that require selecting necessary words in the box are difficult.
– M (1), L (2)
- 7) How about splitting a sentence into several blanks and giving one credit for each blank? – M (1), L (1)
- 8) I think the presentation of the question should be clearer. If the question has a certain condition, it should be salient. – M (2)

9) It is difficult to guess what supply questions would be on the exam. So it is hard to prepare for them. – M (1), L (2)

▪ Comments on the test as a whole

10) Text memorization is necessary to solve the questions. – M (2), L (2)

11) I don't like text alterations or modifications. – M (2), L (1)

12) I got embarrassed when I came across the content that I didn't learn.
– M (2), L (2)

13) I am satisfied with the test. If I study hard, I will be rewarded.
– H (1), M (2)

14) I think the quality of the questions were mostly alright. – H (2), M (1)

15) I think the test was appropriately difficult. I hope the next exam would be like this. – H (3)

16) Compared to the lesson, the test is more difficult. – M (1), L (2)

17) I don't have to rely on private education to prepare for the test because studying the textbook during class was enough to get a high score.
– H (1), M (1)

Although teachers made some changes to make the supply items easier for the students on the final exam, many students made comments on the supply items which showed their dissatisfaction and frustration. Some mid-level and low-level students stated that the supply items were still too difficult for them. It should be

noted that on the final exam, partial credit was given when students made minor spelling or article errors, which was different from the midterm, where no partial credit was allowed. Teachers expected that partial credit would raise scores for the supply items for many students and in turn, their total score and attitudes about the exam would be improved. However, the result was different. Giving partial credit did not raise the score a great deal among the students, and many students mentioned they did not feel that there was much of a difference between the supply items on the midterm and on the final exams. As some students noted above, more partial credit should be given.

Compared to the comments on the midterm exam, the high-level students showed more satisfaction with the test in general. They commented that the difficulty level of the final exam was appropriate, and they were rewarded by the results. Still, the students at the mid-level and low-level groups complained about difficult words that they did not know, text alterations or modifications, and about the difficulty of the test as a whole. On both the midterm and final exams, some mid-level and low-level students responded that it was not easy to answer correctly without textbook memorization.

4.3.3. The Second Feedback Session

The second feedback session was held two weeks after the administration of the final exam. All of the procedures and information provided were similar to those of the midterm exam, but the data were discussed compared to those of the

midterm exam to determine if any significant changes existed.

The teachers' effort to improve their tests did not influence the mean and standard deviation of the test very much. However, the MC items on the final exam improved in terms of the item analysis. This was very rewarding to the teachers who had put a lot of effort to produce a better test. One teacher commented:

I was surprised when I saw the changes in the statistical indices of the final exam. The feedback session did work... If we continue to practice having a feedback session after an exam, the quality of tests will improve remarkably.
(Teacher A)

The teachers attributed this result to their effort to take more time in item writing, especially considering the creation of attractive distracters. For instance, they produced all of the options in English in a content-agreement question on the final exam, which in turn prevented students from easily guessing the answer using their memory. Nonetheless, some of the questions were not 'good' items in terms of test statistics, indicating that efforts regarding these changes should be continued.

The extent to which students considered that the test measured their English ability fairly enough increased slightly on the final exam, but its mean (2.94) remained low. The difficulty of the supply items may have influenced this result, as many students revealed their dissatisfaction and frustration with those items.

Teachers changed their scoring standards to give partial credit for minor errors such as misspellings on the final exam. However, this did not change the total correctness or students' satisfaction levels much, as shown in their comments. Contrary to their expectations, the teachers again failed to control the difficulty and validity of the supply items. While high-proficiency students commented that the items were appropriately difficult, the mid-level and low-proficiency students complained that the items were very difficult in terms of the question form and scoring method. The teachers needed to come up with new ways to improve them. They decided to use a variety of question types rather than sentence-completion on the prospective exams. Furthermore, they were determined to expand the range of partial credit to raise the scores and students' satisfaction levels with the supply questions. For these changes, they could refer to tests from neighborhood schools and materials provided by education offices or workshops.

Compared to the midterm exam, the high-level students revealed their satisfaction with the exam more frequently in their comments. Test fairness on the final exam also improved the greatest in this group. This shows that the teachers' effort to improve their exam affected one specific group more than the other groups. Still, the mid- and low-level proficiency students had big problems in supply items and they complained that they have to memorize texts in the textbook to solve the questions. This demonstrated that the teachers needed to consider group differences more thoroughly and produce items that were fair to most students and that were possible to solve if they studied well.

4.3.4. The Teachers' Views on the Effects of the Feedback Sessions

After the second feedback session, teachers answered questionnaire III in order to identify how they valued the feedback sessions. Their reactions to these meetings were very positive. They had had certain thoughts and beliefs about the exam before, but they realized how the test actually was through the students' feedback and detailed statistics. Some of the assumptions that the teachers had had before the feedback sessions were correct, while some were not. Major remarks regarding the effects of the feedback sessions are presented below.

The feedback sessions

- 1) gave useful information on how to choose the types of items and how to write attractive options
- 2) gave useful information on how students reacted to supply items
- 3) were good opportunities to think about the difficulty, validity, and form appropriacy of each question
- 4) would influence how to teach in class in turn. In order to prepare the students for the exam, more preparation is needed
- 5) gave useful information on how to control the difficulty of the test overall according to the students' level
- 6) reminded us of the importance of careful test writing and of discussing the test among test writers

On the other hand, one teacher pointed out that if the teachers could interview some students immediately after the exam, they could obtain more thorough and detailed information on the exam.

All of the teachers anticipated that the midterm exam for the next semester would improve in terms of its quality. Below is a remark from one of the teachers.

I once participated in the workshop led by an assessment consultant from the local educational office. At that time, an expert who was also a teacher visited the school and consulted on the tests. However, it was not specific and it was mostly on editing problems. I didn't think that the test quality improved after the consultations. However, through our feedback sessions, I could see very well how the test that I had written actually was in various aspects. I believe that it was a lot more meaningful than a general consulting session by an expert who does not know the context of the school and the class. (Teacher C)

4.4. Results of the Statistical Analyses of the Exams

This section attempts to provide results of the statistical analyses of the midterm and final exams to determine if any significant improvement was made. The following three sections outline the result from the analyses: (1) the descriptive statistics of the midterm and final exams, (2) difficulty indices of the two tests, (3) discrimination indices of the two tests, and (4) distracter analysis of the two tests. The last subsection summarizes the major findings of the section.

4.4.1. Descriptive Statistics of the Exam Scores

Table 4.6 shows the scores of the midterm exam and the final exam.

Table 4.6
Descriptive Statistics of the Midterm and Final Exams

Test	Midterm (N=764)		Final (N=767)	
	Mean	SD	Mean	SD
MC items	57.23	12.05	55.92	15.51
Supply items	8.82	6.18	8.52	5.89
Total	66.15	16.57	64.45	19.95

*The full scores of the MC items and the supply items are 80 and 20 respectively.

As Hughes (2003, pp.12-13) stated, the purpose of achievement tests is to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. Therefore, such tests would be expected to have means substantially above 0.5, as the majority of students would be expected to score highly. In this context, the means of the midterm exam and the final exam are considered appropriate for achievement tests, as they were around 0.65. However, when separating the MC items from the supply items, the means of the MC items are rather high while those of the supply items

are comparably low on both tests.

4.4.2. The Difficulty Indices of the Midterm and Final Exams

Item difficulty is one of the most useful indices of an item's quality. Often referred to simply as a p value, it can range from 0 to 1.00, with higher p values indicating items that more examinees answer correctly.

Students' answers to MC questions were analyzed with SPSS to obtain the item difficulty index. Table 4.7 presents the item difficulty indices for the midterm and final exams.

Table 4.7

Item difficulty Indices (p) of the Midterm and Final Exams

Item No.	Midterm	Final	Item No.	Midterm	Final
1	.80	.82	13	.78	.82
2	.76	.87	14	.90	.69
3	.95	.49	15	.63	.87
4	.67	.60	16	.05	.62
5	.96	.36	17	.59	.86
6	.73	.88	18	.93	.74
7	.75	.86	19	.32	.84
8	.56	.80	20	.82	.65

9	.95	.86	21	.26	.67
10	.97	.24	22	.79	.72
11	.97	.62	23	.72	.61
12	.91	.81			

Typically, items with high p values of 0.75 and above are referred to as ‘easy’ items, while items with low p values of 0.25 and below are described as ‘difficult’ items. The acceptable range is often taken as 0.25–0.75 (see Cangelosi 1990). Using this standard, the item evaluation according to the difficulty index for the two tests was carried out as shown in Table 4.8.

Table 4.8

Item Evaluation by the Item Difficulty Index for the Two Tests

Item difficulty index	Item evaluation	No. of items	
		Midterm	Final
below 0.25	difficult	1	1
0.25 – 0.75	moderate	9	11
above 0.75	easy	13	11

Table 4.8 shows that both exams consisted of approximately 50 percent moderate items and 50 percent easy items, also showing that difficult items were extremely rare. Even if these were achievement tests that examined how successfully students learned the content, this imbalance between easy and

difficult items seem excessive.

The number of moderate items increased on the final exam and in turn, the number of easy items decreased. On the midterm exam, too many items had p values that exceeded 0.90 in their item difficulty index, which indicates that these were extremely easy for students, whereas no such items were found on the final exam. Moreover, the item difficulty index of item 16 on the midterm exam was only 0.05, which means that this item was extremely difficult for all students, including high-level students. On the final exam, there was no extreme index.

4.4.3. The Discrimination Indices of the Midterm and Final Exams

Another classical statistic for classifying items as ‘good’ or ‘bad’ is the item discrimination index. By definition, an item discrimination index tells us how frequently an item is answered correctly by those who perform well on the whole test. Ebel (1979) offered the following experience-based guidelines for indicating the quality of test items.

Discrimination index	Item evaluation
.40 and above	Very good items
.30 –.39	Reasonably good but possibly subject to improvement
.20 –.29	Marginal items, usually needing and being subject to improvement
.19 and below	Poor items, to be rejected or improved by revision

In order to obtain discrimination indices of the two tests, students' answers to MC questions were analyzed with SPSS again. Table 4.9 shows the results.

Table 4.9

Item Discrimination Indices (*D*) of the Midterm and Final Exams

Item No.	Midterm	Final	Item No.	Midterm	Final
1	.49	.43	13	.47	.50
2	.34	.46	14	.45	.50
3	.44	.42	15	.48	.48
4	.37	.57	16	.05	.51
5	.46	.35	17	.49	.32
6	.39	.43	18	.41	.55
7	.43	.53	19	.38	.43

8	.32	.22	20	.53	.48
9	.35	.25	21	.39	.56
10	.42	.38	22	.43	.57
11	.39	.51	23	.38	.57
12	.40	.43			

These figures were then categorized according to Ebel's guidelines for evaluating item discrimination. Table 4.10 presents the results.

Table 4.10

Item Evaluation by the Item Discrimination Index for the Two Tests

Item discrimination index	Item evaluation	No. of items	
		Midterm	Final
.40 and above	Very good items	13	18
.30 – .39	Reasonably good items	9	3
.20 – .29	Marginal items	0	2
.19 and below	Poor items	1	0

Table 4.10 shows that most of the items on the two tests could be considered 'good' items with regard to the item discrimination index. The number of 'very good' items significantly increased on the final exam from 13 to 18. At the same time, there was one 'poor' item which definitely failed to discriminate the upper group from the lower group on the midterm exam, but

such an item no longer existed on the final exam. It appears that compared to the midterm exam, the quality of the final exam had improved in terms of discrimination power.

4.4.4. Distracter Analysis of the Midterm and Final Exams

Items that appear to be in need of revision on the basis of their p values or their discrimination indices needed more thorough investigation. For multiple-choice items, distracter analysis in which how the test takers respond to the item's distracters can provide further insight. Table 4.11 presents the percentage of selecting each option and the degree of distracter attractiveness¹⁰ of the midterm exam.

¹⁰ The degree of distracter attractiveness (P_o) was calculated as the percentage of incorrectly selected answers by those who did not answer correctly (Sung, 2005).

$$P_o = 1 - P/Q - 1$$

P_o = The percentage of selecting incorrect answers

P = Item difficulty

Q = The number of options

Table 4.11
Percentage of Selecting Each Option
and Degree of Distracter Attractiveness of the Midterm Exam

Item No.	The percentage of selecting each option					Correct answer	Degree of distracter attractiveness
	A	B	C	D	E		
1	0.80	0.04	0.01	0.11	0.04	A	0.05
2	0.04	0.07	0.05	0.77	0.07	D	0.06
3	0.01	0.00	0.95	0.03	0.01	C	0.01
4	0.04	0.05	0.67	0.05	0.20	C	0.08
5	0.01	0.01	0.01	0.01	0.96	E	0.01
6	0.02	0.07	0.13	0.73	0.06	D	0.07
7	0.04	0.12	0.05	0.75	0.04	D	0.06
8	0.20	0.04	0.04	0.57	0.15	D	0.11
9	0.02	0.01	0.02	0.95	0.01	D	0.01
10	0.03	0.95	0.01	0.01	0.00	B	0.01
11	0.01	0.00	0.01	0.01	0.97	E	0.01
12	0.05	0.01	0.91	0.01	0.02	C	0.02
13	0.02	0.06	0.04	0.10	0.78	E	0.06
14	0.05	0.90	0.03	0.01	0.01	B	0.03
15	0.63	0.10	0.05	0.12	0.09	A	0.09
16	0.05	0.31	0.01	0.45	0.17	A	0.24
17	0.03	0.03	0.59	0.33	0.03	C	0.10
18	0.93	0.02	0.02	0.02	0.01	A	0.02

19	0.28	0.12	0.16	0.13	0.32	E	0.17
20	0.06	0.09	0.82	0.02	0.01	C	0.05
21	0.05	0.26	0.27	0.36	0.06	B	0.19
22	0.01	0.02	0.02	0.16	0.79	E	0.05
23	0.01	0.01	0.08	0.72	0.18	D	0.07

According to Table 4.11, the degree of distracter attractiveness ranged from 0.01 to 0.24. The degree of distracter attractiveness for seven items was very low (under 0.02). While the correct answer for item 16 was option A, half of the students selected D as an answer, and options B and C were also very attractive to the students.

Like on the midterm exam, the percentage of selecting each option and the degree of distracter attractiveness on the final exam were calculated. Table 4.12 presents the percentage of selecting each option and the degree of distracter attractiveness of the final exam.

Table 4.12
Percentage of Selecting Each Option
and Degree of Distracter Attractiveness of the Final Exam

Item No.	The percentage of selecting each option					Correct answer	Degree of distracter attractiveness
	A	B	C	D	E		
1	0.82	0.02	0.06	0.07	0.02	A	0.05
2	0.02	0.03	0.07	0.88	0.01	D	0.03

3	0.28	0.03	0.49	0.11	0.09	C	0.13
4	0.13	0.08	0.15	0.60	0.04	D	0.10
5	0.36	0.36	0.07	0.11	0.10	A	0.16
6	0.05	0.02	0.02	0.03	0.88	E	0.03
7	0.03	0.86	0.06	0.04	0.02	B	0.04
8	0.01	0.01	0.17	0.80	0.02	D	0.05
9	0.02	0.02	0.06	0.04	0.87	E	0.04
10	0.62	0.08	0.24	0.05	0.01	C	0.19
11	0.13	0.05	0.07	0.13	0.62	E	0.10
12	0.16	0.01	0.81	0.02	0.01	C	0.05
13	0.01	0.06	0.10	0.82	0.01	D	0.05
14	0.05	0.07	0.03	0.69	0.17	D	0.08
15	0.03	0.04	0.04	0.87	0.03	D	0.03
16	0.07	0.62	0.03	0.11	0.17	B	0.10
17	0.86	0.02	0.05	0.03	0.04	A	0.04
18	0.12	0.74	0.06	0.04	0.05	B	0.07
19	0.02	0.84	0.06	0.05	0.04	B	0.04
20	0.04	0.02	0.23	0.06	0.65	E	0.09
21	0.08	0.05	0.67	0.11	0.09	C	0.08
22	0.02	0.18	0.72	0.03	0.04	C	0.07
23	0.61	0.15	0.04	0.13	0.06	A	0.10

According to Table 4.12, the degree of distracter attractiveness on the final exam ranged from 0.03 to 0.19. There were no items with a degree of distracter attractiveness under 0.02. This means that for most items, distracters were more attractive than those on the midterm exam. For item 10, distracter A was attractive to the point that many students did not select the correct answer.

Sung (2005) considered a distracter whose percentage of selection is higher than the degree of distracter attractiveness as an attractive distracter. According to this standard, an analysis of attractive and unattractive distracters was carried out later by the teachers during the feedback sessions.

4.4.5. Summary of the Findings

The present section has explored the test results of the midterm and final exams. A summary of the findings is as follows.

- (1) The means of the midterm and final exams were 66.25 and 64.95 respectively. These means were acceptable as achievement tests, but when separating MC items from supply items, the means of the MC items were somewhat high and those of the supply items were comparably low.
- (2) Except for one or two difficult items, about half of the items were 'easy' items and the rest were 'moderate' ones in both tests. The number of moderate items increased on the final exam. Moreover, there were no

extremely easy or difficult items for which the p value exceeded 0.90 or was lower than 0.10 on the final exam. There were eight such items on the midterm exam.

- (3) The number of 'very good' items with regard to item discrimination increased on the final exam. 78% of the items belonged to this category and there was no 'poor' item. On the midterm exam, 56% of the items were 'very good' items, and there was one 'very poor' item.
- (4) The degree of distracter attractiveness ranged from 0.01 to 0.24 on the midterm exam and from 0.03 to 0.19 on the final exam. The changes do not appear great, but this means that seven extremely unattractive distracters (under 0.02) disappeared on the final exam.
- (5) Regarding item difficulty, discrimination, and distracter analysis for the two tests, it appeared that the quality of the final exam improved overall. However, the results showed that many of the items constructed by the teachers were in fact too easy or too difficult on both exams, indicating the necessity for more revision and modification.

CHAPTER 5.

CONCLUSION

The main focus of the present study was on the investigation of the effects of students' feedback on their teachers' testing awareness and practices. This chapter concludes the present study. Section 5.1 summarizes major findings of this study and discusses implications for language pedagogy. Section 5.2 presents the limitations of this study and suggestions for further research.

5.1. Summary of the Research Findings and Implications

This study investigated the effect of students' feedback on English teachers' test preparation and administration. Specifically, the study addressed three research questions: (1) How do students react to the quality of teacher-made tests? (2) How does the students' feedback influence their teachers' awareness and practice on testing? and (3) How do the teacher-made tests improve after the students' feedback? The answers from about 760 tenth-grade high school students for their midterm and final exams were analyzed. Among them, 115 students were selected to write questionnaires for both tests. Four English teachers participated in the study and their comments on the tests and remarks during the feedback sessions were analyzed through qualitative methods.

With respect to Research Question 1 on students' reactions to their teachers' tests, the means of test fairness on the midterm and final exams were 2.60 and

2.94, respectively. These figures show that the students' attitudes as regards test fairness were not positive despite the significant improvement on the final exam. Because students' attitudes about tests strongly influence test performance (Shohamy, 1982) and students' motivation to study, these teachers must exert more effort when constructing 'fair' tests. Meanwhile, perceived item difficulty, validity, and form appropriacy for each item by students gave very useful insight to the teachers when analyzing test items. A few items that students perceived as difficult were analyzed and reviewed. Fortunately, most items were deemed to have appropriate item validity and form appropriacy on both tests. Students' written comments were also very useful for teachers, helping them to make decisions on test writing tasks. Compared to the MC items, students commented more about the supply items, which means that the supply items needed more revision and improvement. Comments from the students varied according to their proficiency level, indicating group differences in terms of their satisfaction levels and complaints on the exams.

With respect to Research Question 2, teachers' testing awareness and practices showed significant changes throughout the study. While reviewing test statistics and the students' feedback in the first feedback session, they attained meaningful insight regarding their test. They discussed the weak points of their test, and how to improve them. Teachers endeavored to apply what they learned during the creation of the final exam. In the second feedback session, the teachers were satisfied with the test results on the final exam, as they saw some improvement on the item analysis and feedback from the students. Still, some of

the questions – especially supply items – required further improvement, indicating that more effort by teachers was needed. Overall, teachers’ attitudes towards feedback sessions were very positive. The feedback sessions made the teachers realize what their tests were like, how to change them, and what results can be brought forth by their efforts. The teachers valued the feedback sessions more than other general assessment workshops because they were voluntary efforts by the teachers who knew the context of the school and class very well.

With respect to Research Question 3, the test results presented some notable findings. First, both tests had acceptable means of around 65.0 as an achievement test, but the means of the MC items were much higher than those of the supply items on both tests. Although the teachers tried to raise students’ scores on the supply items, their effort did not bring the expected results. More effort to find a balance between these two types of items was needed. Second, the item analysis of the MC items on both tests proved that there was some improvement in the quality of the final exam. Regarding item difficulty, discrimination, and distracter analysis, better results were shown on the final exam. This means that students’ feedback on the tests and teachers’ effort to improve their own tests did in fact affect the test quality. Third, many of the items constructed by teachers were found to be too easy or too difficult on both exams, indicating the necessity for more revision and modification by the teachers.

These findings have four major implications. First, the overall quality of teacher-made tests was found to be lacking. There were too many easy items on the school achievement tests produced by teachers, although the number of

extremely easy items decreased significantly on the final exam. Regarding the supply items, the situation was the opposite. They were too difficult for most students and thus their means were around 8.5 out of 20.0 on both tests. This may have arisen because it was the first year that the school achievement tests had to contain supply items according to the local education office. Thus, the teachers who participated in this study were not experienced in writing supply items and were not fully prepared to do so. They reported that they had not participated in any detailed or specific training sessions or workshops regarding the introduction of supply items. This indicates the need for continuous and tailored teacher training on assessment as well as teachers' voluntary efforts to develop themselves as better test writers.

Second, this study enlightened many English teachers to the importance of gaining students' feedback to improve their test quality. In this study, the teachers revealed their surprise at the students' feedback, as some of them were beyond their expectations. Teachers responded that the students' feedback helped them consider and evaluate their test items and improve them on future tests. This study also provided useful means of gathering students' feedback and analyzing test items. The student questionnaires used in this study can also be used by other English teachers who are interested in improving their test writing abilities. In addition to receiving students' feedback, item analyses can also give teachers useful insight into their test items. In this study, item difficulty, discrimination, and distracter analysis were all analyzed with SPSS for a more detailed analysis,

but teachers can refer to the information provided by NEIS for practical matters¹¹.

Third, this study brings about a discussion on the need for teachers to recognize the importance of language testing and to refine testing techniques for better instruction. In many schools, tests end without any discussion about their quality and do not bring any washback for future tests. This should not be the case anymore for better testing and instruction. The importance of tests developed by classroom teachers has become more significant because they make significant decisions about students. For example, school achievement tests can play a very important role when students apply for a university in the future. This is why more attention should be given by teachers to the quality of their examinations. In the present study, although it was a brief opportunity to evaluate test items with students' feedback, the teachers found that the effort was worthwhile. The teachers' efforts to improve test items with the students' feedback raised their awareness and improved their testing practices. These efforts should be continued for better testing and instruction.

Fourth, some comments by the students in this research touched on controversial issues related to assessment in schools. Comments on difficult words that the exam contained, text modification and paraphrasing, and the content they did not learn on the two exams all reflect the students' complaints about the unfairness of the tests. If a school exam is an achievement test, as defined earlier in this study, it should examine what the students have learned. However, this is not

¹¹ NEIS provides teachers with test results, including the percentage of correctness and the item response distribution.

always possible for most teachers because they have to rank the students and thus have to include some difficult questions. It is their duty to strike a balance between the nature of achievement tests and the test's administrative function of ranking students. Similarly, the fact that the students learn in different classes but take the same test poses as to whether the exam is actually assessing what was taught. To solve these problems, assessments in schools need to change from the present materials-based teaching and testing to curriculum-based teaching and testing. Teaching and testing are more connected when students learn to achieve certain objectives and take a test based on them rather than when they learn content in a certain textbook and take a test based on it. Those who are responsible for the teaching and testing English in Korea should consider these issues thoroughly.

5.2. Limitations of the Study and Suggestions for Future Research

This section discusses the limitations of the study briefly and gives suggestions for future research.

First, and most importantly, the present study did not last for an adequate period of time to observe the difference in the test quality brought about by the teachers' discussions of students' feedback. Only two feedback sessions occurred, and only two test results were analyzed and compared during the study. Further study is needed to cover a longer period of time to track the teachers' changes in testing awareness and practices.

Second, this study did not investigate the percentage of correct answers, the difficulty level, or the degree of discrimination of each supply item, although supply items brought many issues throughout the study. The present study employed the entire tenth grade in a school as participants when discussing test results, as their answers were automatically collected through OMR cards and saved on a computer system as a school achievement test. However, the supply items were all scored by individual teachers by hand and the total score of each student was uploaded to a computer by the teachers, to be added to the scores of the MC items. This is why the present study could not analyze the individual supply items. Further study is needed to investigate the supply items more thoroughly, as their importance is growing compared to their poor quality shown in the present study.

Third, this study only explored the quality of test items produced by classroom teachers. Besides test items, there are many aspects of testing that teachers need to consider to construct a good test. For example, they can investigate if the scoring rubric for supply items is appropriate and whether the scoring is reliable. Test reliability depending on the physical context or the procedure of the test can also be examined. Indeed, the quality of the test induces so many other issues to be considered. Thus, it would be useful if future research is conducted on the issues pertaining to the testing such as items as well as rubrics, test methods, rating scales and other facets of test design.

REFERENCES

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test supply and evaluation*. Cambridge, UK: Cambridge University Press.
- Alderson, J. C. (1998). New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing*, 5, 220-32.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing*, 7, 13-30.
- Brown, A. (1993). Test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, 10, 277-303.
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Longman.
- Cangelosi, James. S. (1990). *Designing tests for evaluating student achievement*. New York: Longman.
- Carter, K. (1984). Do teachers understand principles for writing tests? *Journal of Teacher Education*, 35(6), 57-60.
- Chae, J. H. (2004). *An analysis of regular English tests in high schools*. Unpublished master's thesis. Korea University, Seoul, South Korea.
- Choi, I. C. (1997). Essential test method facets of a general English proficiency test and their validity as perceived by test-takers. *Language Research*,

33(4), 773-799.

Choi, I. C. (1999). Test fairness and validity of the TEPS. *Language Research*, 35(4), 571-603.

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, 226-242.

Ebel, R. (1979). *Essentials of educational measurement*. Third ed. Englewood Cliffs, NJ: Prentice-Hall

Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. New York: Macmillan.

Hong, K. O. (2009). *An analysis of regular English tests in middle schools based on the 7th national curriculum*. Unpublished master's thesis. Sookmyung University, Seoul, South Korea.

Hughes, A. (2003). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.

Hutchinson, J., & Waters, A. (1987). *English for specific purposes*. Cambridge, UK: Cambridge University Press.

Im, B. B. (1996). A suggestion for better test supply 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.

Johnson, R., Becker, P., & Olive, F. (1999). Teaching the second-language testing

- course through test development by teachers-in-training. *Teacher Education Quarterly*, 26(3), 71-82.
- Joo, M. J. (2008). Face validity of direct and semi-direct oral tests from test taker perspectives. *A New Korean Journal of English Language and Literature*, 50(2), 285-310.
- 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.
- Jung, M. A. (2004). *An analysis of English tests in high schools*. Unpublished master's thesis. Korea University, Seoul, South Korea.
- Kenyon, D. M., & Stansfield, C. (1991). A method of improving tasks on performance assessments through field testing. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Kim, C. S. (2003). *An analysis of regular English tests in middle schools*. Unpublished master's thesis. Korea University, Seoul, South Korea.
- Kim, J. S. (2005). *An analysis of regular English tests in middle schools with additional suggestions*. Unpublished master's thesis. Chung-Ang University, Seoul, South Korea.
- 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.

- Kirschner, M., Spector-cohen, E., & Wexler, C. (1996). A teacher education workshop on the supply of EFL tests and materials. *TESOL Quarterly*, 30(1), 85-111.
- Ko, J. K. (2007). *An analysis of test items in middle school English examinations*. Unpublished master's thesis. Hanyang University, Seoul, South Korea.
- Madson, H. S. (1983). *Techniques in testing*. New York: Oxford University Press.
- Park, J. H. (2010). *An analysis of the test items in the secondary school English tests*. Unpublished master's thesis. Pusan National University, Pusan, South Korea.
- Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective.*, Englewood Cliffs, NJ: Prentice Hall.
- Popham, W. J. (2000). *Modern educational measurement: practical guidelines for educational leaders*. Boston: Allyn & Bacon.
- Popham, W. J. (2001). *The truth about testing*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2005). *Classroom assessment: what teachers need to know?* Fourth ed. Needham Heights, MA : Allyn & Bacon.
- Reynolds, C. R., Livingston, R. B., & Willison, V. (2006). *Measurement and Assessment in Education*. Boston: Pearson Education.
- Scott, M. L. (1986). Student affective reactions to oral language tests. *Language Testing*, 3, 99-118.
- Shin, D. I., Kim, N. H., & Kang, S. J. (2010). Korean examinees' perceptions of testing spoken English: a movement toward proficiency. *Studies in English*

Language & Literature, 36(2), 263-288.

Shohamy, E. (1982). Affective considerations in language testing. *Modern Language Journal*, 66, 13-17.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning*, 33(4), 527-539.

Soranastapom, S., Chantarasom, A., Suwattananand, N., Javitayanujit, S., & Suwanwongse, S. (2005). The relationship between standardized and teacher-produced tests. *The Journal*, 1, 31-47.

Sung, T. J. (2005). *Modern educational assessment*. Seoul: Hakjisa.

Worthen, B. (1999). *Measurement and assessment in schools*. New York: Allyn & Bacon.

Zeidner, M. (1988). Sociocultural differences in examinees' attitudes toward scholastic ability exams. *Journal of Educational Measurement*, 25, 67-75.

Zeidner, M., & Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as foreign language. *Language Testing*, 5, 100-114.

Zeidner, M. (1990). College students' reactions towards key facets of classroom testing. *Assessment and Evaluation in Higher Education*, 15, 151-69.

APPENDICES

APPENDIX 1-1. The Midterm Exam Paper.....	89
APPENDIX 1-2. The Final Exam Paper.....	95
APPENDIX 2. Questionnaire for the Students	100
APPENDIX 3-1. Questionnaire for the Teachers I	108
APPENDIX 3-2. Questionnaire for the Teachers II.....	110
APPENDIX 3-3. Questionnaire for the Teachers III	111

APPENDIX 1-1. The Midterm Exam Paper

2010학년도 1학기 중간고사 평가 문항지 (1학년)

과목명 : 영 어

계열 : 1 과목코드 : 10 시행일시 : 4월 28일 수요일 1교시 OO고등학교

- 답안지에 필요한 연필 사항(반 번호, 이름 등) 및 과목 코드들 정확히 표기한 후 답안을 작성하십시오.
- 객관식 : 문수에 맞는 답을 글라 답안지의 해당 번호에 바르게 표기하십시오.
- 서술형 : 답안지 뒷면에 정확히 기입하십시오. 서술형 답안은 경색이나 검정색 펜으로 작성하십시오.
- 배점은 각 문항 끝에 표시하였으므로 참고하십시오.

1. 다음 대화의 밑줄 친 (A), (B)에 들어갈 말을 바르게 짝지은 것은? (3.4점)

A: Mr. Adams, do you mind if I borrow a book?
 B: (A), Sujin. Which one do you want to read this time?
 A: I'd like to try fantasy novel today.
 B: Oh, (B). Let's see.. Oh, here's a good one.
 A: Thank you very much, Mr. Adams.

- | | |
|-----------------|-----------------------------------|
| (A) | (B) |
| ① Of course not | that sounds interesting |
| ② Of course not | I'm afraid I can't |
| ③ Yes, I do | thank you for your kindness |
| ④ Yes, I do | that is a good idea |
| ⑤ Never mind | I'm grateful for your help |

2. 다음 짝지어진 대화가 어색한 것은? (3.8점)

- ① A: We must prepare for the garage sale till tonight.
 B: Oh gosh, but I don't want to do that!
 It's so demanding.
- ② A: Why don't we watch the musical, Miss Saigon?
 It'll be awesome and we won't forget tonight forever!
 B: Don't you remember we got a conference at 7 o'clock?
- ③ A: Whatever people say to me, I think money is the most valuable thing in the world. It lasts forever!
 B: I wish you wouldn't think that way.
 Money is valueless when you happen to lose your health.
- ④ A: What do you think of eating pizza for dinner with me?
 B: Hmm, I think it's a good idea. It really hurts my stomach.
- ⑤ A: I'm upset. I'm not happy about my brother's behavior!
 I'm going to beat him to death!
 B: Don't you remember father's last warning?
 If it happens again, you would be in a big trouble.

3. 다음 대화의 빈 칸에 들어갈 말을 보기에서 골라 바르게 배열한 것은? (3.2점)

W: Jinsu, how do you like your new high school?
 M: It's great! The school is big and beautiful.
 The teachers are very kind and friendly, too.
 W: (A) Do you have any plans for the new school year?
 M: Yes, (B)
 W: Good for you.
 M: But I don't know how to improve my English. Do you have any advice?
 W: (C) You'll be able to practice English a lot there.
 M: That's a good idea.

< 보기 >

- ㄱ. I'm sorry.
- ㄴ. I'm glad to hear that.
- ㄷ. I've decided to study English a lot.
- ㄹ. Why don't you join the school English newspaper club?

(A) (B) (C)

- ① ㄱ ㄴ ㄷ
- ② ㄱ ㄴ ㄹ
- ③ ㄴ ㄷ ㄹ
- ④ ㄴ ㄷ ㄱ
- ⑤ ㄱ ㄷ ㄹ

4. 다음 글의 빈칸에 들어갈 말로 가장 적절한 것은? (3.4점)

These days, quickness is important even in food. Fast food is popular because it is quick, cheap, and easily available. However, fast food may cause certain health problems because it is high in fat, salt, and sugar.

More recently, many people have been emphasizing slowness in food as an alternative to fast food. One of these people is an Italian writer named Carlo Petrini. He had an idea that he called Slow Food. His concept of Slow Food was to slow down the pace of cooking and eating food. Petrini also felt that food should be produced in a way that _____

- ① can provide more profits to manufacturers
- ② is indifferent to the environmental problems
- ③ is harmless to the environment or human health
- ④ doesn't appreciate the giving of the environment
- ⑤ doesn't improve the environment or human health

[11-12] 다음 글을 읽고 물음에 답하시오.

Sojin: _____ (A) _____ ?
 Hyori: I already feel that high school is different _____ ㉠ _____ middle school. There are more classes and more assignments. I also feel _____ ㉡ _____ more responsible for myself.
 Sora: It is more difficult to make new friends in high school. I think that it'll take some time to become friends with my classmates.
 Yena: My teachers and parents treat me _____ ㉢ _____ a more independent person. I think that I should take care of myself and trust myself more.

11. 위 글의 빈 칸 (A) 에 들어갈 질문으로 가장 적절한 것은? (3.2점)

- ① What do you want to be?
- ② Why don't you introduce yourselves?
- ③ What are you doing to achieve your dream?
- ④ What is your plan for the new school year?
- ⑤ What do you think about your high school life?

12. 위 글의 빈 칸 ㉠, ㉡, ㉢에 들어갈 단어가 바르게 짝지어진 것은? (3.5점)

- | ㉠ | | ㉡ | | ㉢ |
|--------|-------|------|-------|--------|
| ① of | | much | | like |
| ② of | | very | | alike |
| ③ from | | much | | like |
| ④ from | | far | | likely |
| ⑤ from | | very | | likely |

[13-14] 다음 글을 읽고 물음에 답하시오.

(A) What do you think of when you hear the word 'fast'? How about the word 'slow'? Which word best ㉠ describes/prescribes your life? An old saying goes, "Time is money," and this is still true today. These days quickness is considered important in everyday life.
 (B) For example, many people enjoy using computers, cellphones, or other devices. However, when they cannot use these things, people often feel lost.
 (C) Despite these benefits, quickness does have some drawbacks. Some ㉡ advocates/critics complain that people rely too heavily on technology in order to maintain a fast life.
 (D) It is true that quickness can save time and make people's lives easier and more convenient. For example, thanks to speedier ㉢ mean/means of transportation, people are able to save much time when traveling long distances.

13. 위 글 (A)에 이어질 내용을 순서에 맞게 배열한 것으로 가장 적절한 것은? (3.4점)

- ① (B)-(C)-(D)
- ② (B)-(D)-(C)
- ③ (C)-(B)-(D)
- ④ (D)-(B)-(C)
- ⑤ (D)-(C)-(B)

14. 위 글의 ㉠, ㉡, ㉢ 각 네모 안에서 문맥에 맞는 낱말로 가장 적절한 것은? (3.5점)

- | ㉠ | | ㉡ | | ㉢ |
|--------------|-------|-----------|-------|-------|
| ① describes | | critics | | mean |
| ② describes | | critics | | means |
| ③ describes | | advocates | | means |
| ④ prescribes | | critics | | mean |
| ⑤ prescribes | | advocates | | means |

[15-16] 다음 글을 읽고 물음에 답하시오.

Cathy and Jane were good friends. One day, they were walking through the desert together. At some point, they had a big argument, and Cathy insulted Jane. Jane felt bad, but, without **ⓐ** said anything, she wrote in the sand:

Today my best friend insulted me.

The two friends continued **ⓑ** walking, until they found a deep well. The hot desert sun had made them thirsty, so they decided **ⓒ** to stop for a drink. As Jane leaned over the edge of the well, she slipped and fell in. Since Jane couldn't get out by herself, Cathy had to help her. After Jane was saved, she said **ⓓ** nothing, but she wrote on a stone.

Today my best friend saved my life.

Cathy was curious, so she asked, "After I insulted you, you wrote in the sand, but this time you write on a stone. Why?" Jane replied, "When a friend hurts us, we should write it in the sand so that the winds of forgiveness can erase it. But when a friend does **ⓔ** something good for us, we must write it in stone so that no wind can ever erase it."

15. 위 글의 밑줄 친 **ⓐ~ⓔ** 중 어법상 틀린 것은? (3.4점)

- ① ⓐ ② ⓑ ③ ⓒ ④ ⓓ ⑤ ⓔ

16. 위 글의 내용을 한 문장으로 요약하고자 한다. 빈칸 (A)와 (B)에 들어갈 말로 가장 적절한 것은? (3.8점)

When your best friend insults you, you should **(A)** it as soon as possible, whereas you should **(B)** it when you get help from them.

- | (A) | (B) |
|-------------|----------------|
| ① ignore | keep |
| ② accept | remember |
| ③ disregard | promise |
| ④ forgive | pay back |
| ⑤ forget | refuse |

[17-18, 서술형] 다음 글을 읽고 물음에 답하시오.

Bill and Tom grew up together in a small town. **(A)** Bill was strong and brave, Tom was weak and quiet. When other boys tried to make fun of Tom, Bill was always there to help him. Whenever Bill helped him, Tom always smiled and said, "I just felt you would come."

Years later, they served in the army together. When they were in the army, a war broke out. One day, Bill was shot **(B)** battle. Tom wanted to run to Bill and help him. However, the general said, "Bill is probably dead. There's no use risking your life, too." For Tom, his friend's life was more important than anything else. He asked the general again. Finally, the general let Tom go and help Bill. Tom ran as fast as he could. He picked Bill up and carried him on his shoulders to the hospital. But it was too late. Bill had already died. "See, I told you," said the general. "You risked your life for nothing!" "No, I didn't," replied Tom.

"**(C)**, and I have my reward. When I picked Bill up, he said, 'Tom, I knew you would come; I just felt you would come. You're a true friend.'"

17. 위 글의 내용과 일치하는 것은? (3.8점)

- ① Bill and Tom were good friends although they sometimes had an argument.
 ② Bill died because he was shot by the general.
 ③ Tom went to Bill after he got a permission from the general.
 ④ Tom thought Bill was as important as other friends to him.
 ⑤ There was no hospital to take Bill and Bill didn't say anything before he died.

18. 위 글의 빈칸(A), (B)에 들어갈 말을 바르게 짝지은 것은? (3.4점)

- | (A) | (B) |
|-----------|---------------|
| ① While | during |
| ② Since | for |
| ③ For | in |
| ④ Whereas | while |
| ⑤ Though | despite |

[서술형] 위 글의 빈칸 (C)에 들어갈 말을 관계대명사 **what**를 사용하여 다음의 우리말 해석에 맞게 영작하시오. (4.0점)

해석 : 나는 Bill이 원한 것을 했다.

[19-21, 서술형2] 다음 글을 읽고 물음에 답하시오.

Petrini's concept of Slow Food is so popular that it has become the Slow Movement @that includes Slow Cities, Slow Travel, Slow Books, and Slow Money.

The concept of Slow Cities encourages people to slow down and enjoy life in a small city. People in slow cities live a more comfortable life with less traffic, less noise, and fewer crowds. Currently, only cities with a population of less than 50,000 can be called slow cities. (A) _____ in the future, larger cities can also be called slow cities @that they find ways to encourage slow living.

Slow Travel offers an alternative to fast travel. Fast travelers usually move quickly from one place to the next and spend very @little time enjoying their surroundings. (B) _____, slow travelers stay in one place for a long time. They walk and ride bicycles @to enjoy the natural beauty of the surroundings. They also try to participate in local activities and festivals for a chance to meet local people.

The Slow Books Movement _____ (C) _____. These days few people spend time reading for enjoyment. Slow Books encourages people to make the time to enjoy reading for pleasure. It shows people that reading helps them become more @relaxed.

The latest addition to the Slow Movement is Slow Money. The concept of Slow Money is simple. It is currency that is designed by local artists and that can only be used in the local community. This type of money allows customers to contribute directly to their local businesses and share the spirit of community among neighbors.

19. 위 글의 주제로 가장 적절한 것은? (3.8점)
- ① the way to participate in Slow Movement
 - ② the importance of classifying Slow Movements
 - ③ developments of people's concept toward slowness
 - ④ the preference of Slow Food to other Slow Movements
 - ⑤ the spread of Slow Movement including various categories

20. 위 글의 빈칸 (A), (B)에 들어갈 말로 가장 적절한 것은? (3.8점)

- | | | |
|-----------------|-----|-------------------|
| (A) | ... | (B) |
| ① But | ... | In addition |
| ② Whereas | ... | However |
| ③ But | ... | On the other hand |
| ④ Whereas | ... | In addition |
| ⑤ Unfortunately | ... | Nevertheless |

21. 위 글의 밑줄 친 ㉠~㉤ 중 어법상 틀린 것은? (3.8점)

① ㉠ ② ㉡ ③ ㉢ ④ ㉣ ⑤ ㉤

[서술형2] 위 글의 빈칸 (C)에 들어갈 말을 다음의 우리말 해석에 맞게 <보기>에서 **알맞은 단어들** 골라 순서대로 쓰시오. (4.0점)

해석: 슬로우 북 운동은 사람들에게 현대 생활의 빠른 속도를 줄일 것을 요청한다.

< 보기 >

asks, modern life, people, to, reduce, reducing, the fast pace, of, be

[22, 서술형3] 다음 글을 읽고 물음에 답하시오.

Dasom: Do you have any special things that you want to try in high school?

Dahye: I want to make a lot of Korean friends this year. I also want to learn more about Korea and Korean culture.

Jinhee: _____ (A) _____. Then, both of us could understand each other better. It would be an interesting experience.

Sujeong: I want to write some short stories for children. I also want to have them published some day so that children can enjoy my stories.

Somin: I want to work in the school TV station like you. I think I can work as an engineer. I am good with machines and cameras.

Dasom: Thank you for meeting with me today. This interview will help your friends understand more about you. Good luck with the new semester.

All: Thank you very much for having us. It was a good opportunity for us to think about ourselves and our high school lives.

22. 위 글의 내용과 일치하지 않는 것은? (3.2점)

- ① Dahye는 올해 많은 한국 친구들을 사귀고 싶어 한다.
- ② Dahye는 한국과 한국 문화에 대해 좀 더 알기를 원한다.
- ③ Sujeong은 아이들을 위해 짧은 이야기를 쓰고 싶어 한다.
- ④ Somin은 기계를 잘 다루므로 학교 방송국에서 엔지니어를 하고 싶어한다.
- ⑤ 학생들은 인터뷰를 통해 스스로를 더 믿고 돌보는 기회가 되었다고 생각한다.

[서술형 3] 위 글의 빈칸 (A)에 들어갈 말을 다음의 우리말 해석에 맞게 <보기>에 주어진 단어들 모두 **한번씩 사용하여** 영작하시오. (4.0점)

해석: 나는 하루 동안 나의 선생님과 역할을 바꾸어 보고 싶어요.

< 보기 >

I, I, change, wish, with, my, a, teacher, roles, for, could, day

[23, 서술형4] 다음 글을 읽고 물음에 답하시오.

Minji: What is your plan for the new school year?
Eujin: _____ (A) . I hope to visit local historical places at least once a month. I think I can learn more about the history and culture of my town.
Jaemin: First of all, I want to study hard. ㉠ Get/Getting good grades is really important to me. I also want to join ㉡ few/a few clubs. This year, I'd like to join a volunteer club to help others.
Yerin: I'd like to improve my computer skills. Because I'm not good with computers, it usually takes a long time for me ㉢ finishing/to finish my school projects.

23. 위 글의 ㉠, ㉡, ㉢ 각 네모 안에서 어법에 맞는 표현으로 가장 적절한 것은? (3,8점)

- | | | | | |
|-----------|-------|-------|-------|-----------|
| ㉠ | ㉡ | ㉢ | | |
| ① Get | | few | | finishing |
| ② Get | | a few | | finishing |
| ③ Getting | | few | | to finish |
| ④ Getting | | a few | | to finish |
| ⑤ Getting | | a few | | finishing |

[서술형4] 위 글의 빈칸 (A)에 들어갈 말을 괄호 안에 주어진 단어를 사용하여 다음의 우리말 해석에 맞게 영작하시오. (4.0점)

해석: 나에게 가장 중요한 계획은 여행(traveling)입니다.

[서술형5] 다음 글의 빈칸 ㉠, ㉡에 들어갈 말을 주어진 우리말 해석에 맞게 각각 영작하시오. (4.0점)

(A) Friendships need _____ ㉠ _____ garden flowers. Good friendships involve many things: careful listening, forgiveness, honesty, patience, respect, sacrifice, and trust. Without these things, (B) friendships are _____ ㉡ _____ glass. However, with hard work and effort, friendships can become stronger than stone and can last for all time.

- (A) 우정은 정원의 꽃들만큼이나 많은 보살핌을 필요로 한다.
(B) 우정은 유리만큼 약하다.

- 끝 -

APPENDIX 1-2. The Final Exam Paper

2010학년도 1학기 기말고사 평가 문항지 OO고등학교 (1)학년 과목명 : 영어 계열 : 1 과목코드 : 10 시행일시 : 7월 6일 화요일 3교시	결 계	담당	부장	교감	교장
<p> ○ 답안지에 필요한 인적 사항(반, 번호, 이름 등) 및 과목 코드를 정확히 표기한 후 답안을 작성하시오. ○ 재판식 : 문음에 맞는 답을 골라 답안지의 해당 번호에 바르게 표기하시오. ○ 서술형 : 서술형 답안지에 정확히 기입하시오. ○ 배점은 각 문항 끝에 표시하였으므로 참고하시오. </p>	<p>3. 다음 글의 빈칸에 공룡으로 들어갈 말로 가장 적절한 것은? (3.4점)</p> <p style="text-align: center;"> Let's find out how. Most cell phones use radio waves. A radio in a car also uses radio waves. When a car travels far away from a city, it becomes impossible to hear the radio. The reason is that the radio is too far away from the signal tower. The radio waves are _____. The same thing happens to cell phones. They need to be close to a signal tower to receive radio waves. If you took your cell phone into space, you would be too far away from the nearest signal tower. The radio waves would be _____. If someone built a signal tower in space, however, then yes, your cell phone could work. </p> <p> ① too far to send ② too indirect to send ③ not strong enough to receive ④ not straight enough to receive ⑤ not observable enough to signal </p>				
<p>1. 다음 짝지어진 대화가 어색한 것은? (3.2점)</p> <p>① A: Have you ever been to New Zealand? B: Yes, I have never been to any foreign country.</p> <p>② A: Will you show me how to get to the ABC mart? B: Sure, just go straight following this street.</p> <p>③ A: Do you know how to play tennis? B: No, I don't. I don't enjoy exercising at all.</p> <p>④ A: Do you know anything about the rumor? B: Not at all, I've never heard about it.</p> <p>⑤ A: It's going to rain soon. Make sure you bring your umbrella. B: OK, I will bring it.</p>	<p>4. 다음 글에서 전체 흐름과 관계 없는 문장은? (3.6점)</p> <p style="text-align: center;"> The world is filled with wonders. They are everywhere we look and in everything we do. Even the most ordinary things have scientific principles behind them. ①You do not have to be a scientist to find them. ②All you have to do is look and wonder. ③This is the basis of science. ④Scientific principles cannot explain the wonders of our surroundings. ⑤Observing, comparing, recording, and sharing information are practical scientific skills that people can use every day. What will you do to find the science around you? </p>				
<p>2. 다음 대화의 빈 칸 (A), (B), (C)에 들어갈 말을 보기에서 골라 적절하게 배열한 것은? (3.2점)</p> <p>Andy : What did you think of the pictures in the art gallery? Sujin : The pictures were nice, but some of the people there really bothered me. Andy : What do you mean? Sujin : Those people on their phones were so noisy. _____ (A)</p> <p>Andy : You're right. Sujin : _____ (B) turn off their phones in an art gallery. Andy : I agree. They should at least go outside to talk. Sujin : _____ (C)</p>	<p>5. 다음 글에 드러난 필자의 어조로 가장 적절한 것은? (3.4점)</p> <p style="text-align: center;"> I thought good cellphone etiquette was common sense, but I was wrong. Some people never turn off their phones, not even during concerts or performances. They don't even set them to vibrate. Also, when people get a call on a bus or train, they talk so loudly. Why don't they think of the other people around them? </p> <p> ① critical ② ironical ③ neutral ④ optimistic ⑤ sentimental </p>				
<p style="text-align: center;">< 보기 ></p> <p> 가. I beg your pardon? 나. Don't you think so? 다. I had no idea how to 라. People are supposed to 마. That's a good point. 바. I'm glad to hear that. </p> <p> (A) (B) (C) ① 가 나 다 ② 가 나 다 ③ 나 다 라 ④ 나 라 마 ⑤ 가 다 바 </p>	<p style="text-align: center;">2010학년도 1학기 기말고사 1학년 영어 1/5</p>				

6. 다음 글의 바로 앞에 올 내용으로 가장 적절한 것은? (3.0점)

However, most children are naturally curious about the world around them. They ask questions like, "Why is the sky blue?, How do roller coasters work?, or Why do some people get headaches when they eat ice cream?" The only way to answer these questions is to use science. Science explains the wonders of the world around us. As the famous physicist, Michael Faraday said, "To have a scientific attitude toward the world, one must be able to imagine wonderful things - even things that are too wonderful to be true."

- ㉠ 아이들이 가지는 호기심의 근원
- ㉡ 과학법칙을 실생활에 적용하는 것의 한계
- ㉢ 어른들의 질문과 어린이들의 질문의 차이점
- ㉣ 어린이들의 주변 환경에 대한 비범한 관찰력
- ㉤ 모든 주변 현상을 당연히 여기는 어른들의 시각

7. 주어진 글 다음에 이어질 글의 순서로 가장 적절한 것은? (3.4점)

Etiquette is more than just rules about what to do and what not to do.

- (A) Since etiquette forms the basis for getting along with others, it is important everywhere you go, even on the Internet.
- (B) It is a concept about showing concern for the people around us.
- (C) Here are some examples of online and offline etiquette around the world.

- ㉠ (A)-(C)-(B) ㉡ (B)-(A)-(C) ㉢ (B)-(C)-(A)
- ㉣ (C)-(A)-(B) ㉤ (C)-(B)-(A)

8. 다음 글의 주제로 가장 적절한 것은? (3.8점)

Good money management is important for everyone. It involves saving, planning, and smart spending. Even if you have just a little money, you still need to manage it wisely. When you do so, you will be able to reach your goals and discover the joy that comes from good money management. It may not be easy at first, but if you take the time to make saving, planning, and smart spending a habit, it really will be as easy as ABC. All you have to do is set your aim, make your budget, and collect your coins and currency!

- ㉠ some useful advice on how to shop
- ㉡ how to make money in a short time
- ㉢ the best way to achieve saving goals
- ㉣ the importance of good money management
- ㉤ do's and don'ts for good money management

9. 다음은 민호의 예산안이다. 도표의 내용과 일치하지 않는 것은? (3.8점)

<Minho's Budget>		
★ Buy things after looking at three different types.		
★ Always put your change in a piggy bank.		
Monthly Income	My allowance	₩ 30,000
	Gift money	₩ 10,000
	Total	₩ 40,000
Monthly Spending	Entertainment	₩ 10,000
	Snacks	₩ 15,000
	Others	₩ 5,000
	Total	₩ 30,000
Monthly Savings		
Total monthly savings: ₩10,000		
*I'll spend ₩10,000 less every month and put it into my savings account. I can buy an MP3 player in about ten months.		

- ㉠ Minho tries to compare three different types when purchasing.
- ㉡ Minho has a place to put his changes.
- ㉢ Minho's earning is 10,000 won higher than his expense.
- ㉣ Minho can save 10,000 won every month.
- ㉤ Minho wants to have an MP3 player which is about 200,000 won.

[10-11] 다음 글을 읽고 물음에 답하십시오.

A ring that can measure a person's mood? It may ① sound like science fiction, but actually it's a scientific fact. Let's find out why. Mood rings are filled with a special material that is ② neither liquid, solid, nor gas. It is the same material that is found in digital watches and calculators. It's called liquid crystal.

One of the special qualities of liquid crystal ③ is that it changes color when the temperature ④ changes. When people feel happy, their body temperature increases slightly. This small temperature change causes the liquid crystal in the mood ring ⑤ turn blue. And when people are stressed, their body temperature decreases slightly and the mood ring turns gray.

10. 위 글의 내용과 일치하지 않는 것은? (3.8점)

- ㉠ Mood rings are not based on fiction but a certain fact.
- ㉡ Calculators and mood rings use the same material.
- ㉢ Color change of liquid crystal causes temperature changes.
- ㉣ Feeling happiness makes the ring turn blue.
- ㉤ When stressed, the rings would turn gray.

11. 위 글의 밑줄 친 ㉠~㉤ 중 어법상 틀린 것은? (3.6점)

- ㉠ ㉠ ㉡ ㉡ ㉢ ㉢ ㉣ ㉣ ㉤ ㉤

[12~13] 다음 글을 읽고 물음에 답하시오.

Last week our school surveyed the spending and saving habits of students to understand how they manage their money. About 300 students took part in the survey, and the results were interesting.
 About 70 % of the students spent between 7,000 and 13,000 won per week. The majority of these students (92%) said that they got their money from their parents. Regarding spending, the students spent most of their money on fast food and snacks, spending time with their friends, transportation, clothes and shoes, and other things.
 Surprisingly, most of the students (92 %) reported that they didn't save any of their money. They said that they had never made a budget. The overall results show that the students don't generally know how to manage their money. Therefore, _____ (A).

12. 위 글의 빈칸 (A)에 들어갈 내용으로 가장 적절한 것은? (3.8점)
- ㉠ students must have known how to manage their money
 - ㉡ it is a good idea to get a part time job to prepare for their future
 - ㉢ students need more information and education about managing their money
 - ㉣ students had better ask their friends to help them as soon as they run out of money
 - ㉤ students ought to make a campaign to get more allowance from their parents

13. 위 글의 제목으로 가장 적절한 것은? (3.2점)
- ㉠ How to be Rich
 - ㉡ Why We Save Money
 - ㉢ Smart Spending Habit
 - ㉣ Students Save Nothing
 - ㉤ The life of Smart Students

[14-15] 다음 글을 읽고 물음에 답하시오.

Nicole Ross ▶ A Student from Australia
 For me, good etiquette means keeping your voice down in the library. If people are talking around me, it interrupts my studying. When I'm in the library, I exercise proper etiquette by not ㉠making too much noise even when turning the pages of my book or ㉡typing on my computer.

Susan Travis ▶ A Student from Canada
 I was really ㉢worried when I got a strange e-mail from my friend the other day. I thought he was angry because HE ㉣HAS TYPED EVERYTHING IN (A)CAPITAL LETTERS. I guess he didn't know that ㉤using all capital letters online is the same thing as yelling offline. He should be more careful to follow the rules of netiquette.

14. 위 글의 밑줄 친 ㉠ ~ ㉤ 중 어법상 틀린 것은? (3.6점)
- ㉠ ㉡ ㉢ ㉣ ㉤ ㉥

15. 아래에 주어진 사전 뜻들이 가운데, 위 글의 밑줄 친 (A)CAPITAL의 의미로 가장 적절한 것은? (3.0점)

capital *n* ㉠ 1. the top part of a column : *He bumped his head on the capital.* ㉡ 2. involving punishment by death : *He committed a capital offense.* ㉢ 3. a large amount of money that is invested or is used to start a business : *We set up a business with a capital of \$ 100,000.* ㉣ 4. having the form and size used at the beginning of a sentence or a name : *Please write in capital letters.* ㉤ 5. the most important town or city of a country, usually where the central government operates from : *Cairo is the capital of Egypt.*

- ㉠ 1. ㉡ 2. ㉢ 3. ㉣ 4. ㉤ 5.

[서술형 II] 다음 글의 빈칸 (A)에 들어갈 말을 <보기>의 단어를 모두 한 번씩 사용하여 다음의 우리말 해석에 맞게 영작하시오. (4.0점/부분점수 있음)

Almost everyone has a story about eating ice cream on a hot summer day and then suddenly getting a very bad headache. Why does this happen? According to scientists, people get ice cream headaches, or 'brain freeze' as it is often called, when _____ (A) _____ touches the roof of the mouth. They think it's because the cold ice cream makes blood vessels in the brain get smaller. This is what causes the pain.

해석: 아이스크림처럼 차가운 무언가

<보 기>
 ice cream, cold, something, like

[16-17, 서술형2] 다음 글을 읽고 물음에 답하십시오.

Everyone knows how to spend money, but not everyone knows how to save it. As an old saying goes, you cannot pick an apple without first planting an apple tree. In the same way, you cannot pick the fruits of wealth without planting a savings tree first. Saving is a good habit (A) . Here are some tips about how to save money.

"A good way to save money is by saving coins. For example, when I buy an ice cream and get a coin back, I put it into a savings jar. I can save a lot of money by saving coins." - Minho

"I keep my money in a piggy bank. Then, I put a picture of what I really want to buy on it . Every time I reach for the money, I see the picture and remember my saving goal." - Sora

"I have two empty cans. I cut a hole in the top of each of them and labeled one "spend" and the other "save." I put half my allowance and gift money in the "spend" can and half in the "save" can. I never empty the "save" can until it is full." - David

16. 위 글의 내용과 일치하지 않는 것은? (3.8점)
- ① Minho has his own special way to save money.
 - ② Minho thinks that saving coins without spending anything can be a good way to manage money.
 - ③ Sora tries to remind herself of her saving goal using a picture.
 - ④ David splits his income evenly and puts them in different cans.
 - ⑤ David doesn't reach for the money in the "save" can until it is filled up.

17. 위 글의 내용을 한 문장으로 요약하고자 한다. 빈칸 ㉠, ㉡, ㉢에 들어갈 말로 가장 적절한 것은? (3.6점)

Students can ㉠ money ㉡ when they make it ㉢ .

- | | | |
|-----------|---------------------|--------------------|
| ㉠ | ㉡ | ㉢ |
| ① save | wisely | a habit |
| ② raise | smartly | a rule |
| ③ spend | technically | an aim |
| ④ collect | inefficiently | an objective |
| ⑤ manage | irrationally | a dream |

[서술형 2] 위 글의 빈칸 (A)에 들어갈 말을 <보기>에서 알맞은 단어를 골라 다음의 우리말 해석에 맞게 영작하십시오. (4.0점/부분점수 있음)

해석: 저축은 절박시킴기에 시간과 노력이 필요한 좋은 습관입니다.

< 보 기 >

what, which, take, takes, effort, to establish, and, time

[18-19, 서술형3] 다음 글을 읽고 물음에 답하십시오.

"..... (A) , but I'm still surprised when students don't bring their books and study materials to class. ㉠ problem is cellphones. Students send and receive text messages during class as if they didn't even know it was bad etiquette to do so." -Salvadoro Garcias

"My hobby is homepage design. I like to post samples of my work on my homepage. The ㉡ day while I was surfing the Internet, I found some of my pictures on another site. They had used my property without my permission. That's not only bad netiquette, but it's also illegal! I had to remind them that netiquette as well as the law requires people to ask for permission to use images and text from other people's sites." -Choe Sangwon

18. 위 글의 내용과 일치하지 않는 것은? (3.8점)
- ① Not all students come to class with their study materials.
 - ② Students send text messages during class because they weren't told not to do so.
 - ③ Sangwon enjoys putting his work on his homepage.
 - ④ Sangwon thinks that his pictures are his property.
 - ⑤ People need to ask for permission when using others' property.

19. 위 글의 빈칸 ㉠, ㉡에 들어갈 말을 바르게 짝지은 것은? (3.4점)

- | | |
|-----------|---------------|
| ㉠ | ㉡ |
| ① Another | another |
| ② Another | other |
| ③ Other | another |
| ④ Other | other |
| ⑤ Others | other |

[서술형 3] 위 글의 빈칸 (A)에 들어갈 말을 현재완료진행 시제를 사용하여 다음의 우리말 해석에 맞게 영작하십시오. (4.0점/부분점수 있음)

해석: 저는 20년 동안 가르쳐 오고 있습니다.

[20-21, 서술형 4] 다음 글을 읽고 물음에 답하시오.

David Foster ▶ A Student from England
 Last week, I missed an important e-mail because my inbox was full. It was full because so many people had forwarded me a bunch of ㉠ annoy / annoying messages. Forwarding jokes, chain letters, or other unimportant e-mails can fill up people's inboxes and ㉡ cause / causing them to miss important e-mail. It can also be dangerous because those types of e-mail messages can contain viruses. Good netiquette means not ㉢ forward / forwarding spam mail to others.
 Every society has rules of etiquette that people are expected to follow. (A) Etiquette covers how and where it is said and done as well as what we say and do. These rules are not simply do's and don'ts, but also concern for others. The first step to good etiquette is putting ourselves in someone else's shoes. (B) _____, we should do to others as we want them to do to us. When we follow proper etiquette, our world, both online and offline, becomes a better place to live in.

20. 위 글의 ㉠, ㉡, ㉢ 에서 어법에 맞는 낱말로 가장 적절한 것은? (3.6점)

- | | | |
|------------|---------|------------|
| ㉠ | ㉡ | ㉢ |
| ① annoy | cause | forward |
| ② annoy | causing | forward |
| ③ annoying | cause | forward |
| ④ annoying | causing | forwarding |
| ⑤ annoying | cause | forwarding |

21. 위 글의 빈칸 (B)에 들어갈 연결어로 가장 적절한 것은? (3.2점)

- | | |
|---------------------|---------------|
| ① Instead | ② Recently |
| ③ In other words | ④ For example |
| ⑤ On the other hand | |

[서술형 4] 밑줄 친 (A)과 같은 뜻이 되도록 **not only**와 **but also** 를 사용하여 영작하시오. (4.0점/부분점수 있음)

(A) Etiquette covers how and where it is said and done as well as what we say and do.

→ _____

[22-23, 서술형 5] 다음 글을 읽고 물음에 답하시오.

(A) At the top of the first hill, a roller coaster is full of potential energy. The higher the hill, the greater the potential energy. Gravity pulls the roller coaster ㉠ up the hill, and _____. This kinetic energy, not an engine, then ㉡ pushes the roller coaster up all the other hills.
 (B) Some roller coasters can travel faster than cars. Much faster! But what makes it possible? You may not know it, but roller coasters do not have engines. Instead, they rely on the scientific principles of gravity, potential energy, and kinetic energy.
 (C) Gravity is the force that pulls everything toward the ㉢ center of the Earth. Potential energy is the power inside an object when it ㉣ changes its position. Kinetic energy is the power which an object gets from ㉤ moving.

22. 위 글의 순서로 알맞은 것은? (3.4점)

- ① (A)-(C)-(B) ② (B)-(A)-(C) ③ (B)-(C)-(A)
 ④ (C)-(A)-(B) ⑤ (C)-(B)-(A)

23. 위 글의 밑줄 친 ㉠~㉤ 중 문맥상 어색한 어휘는? (3.4점)

- ① ㉠ ② ㉡ ③ ㉢ ④ ㉣ ⑤ ㉤

[서술형 5] 위 글의 빈칸에 들어갈 말을 다음의 우리말 해석에 맞게 <보기>에서 알맞은 단어를 골라 순서대로 쓰시오. (4.0점/부분점수 있음)

해석: 위치 에너지는 운동 에너지로 전환된다.

_____ <보기>
 kinetic energy, the potential energy, into, change, changed, is, be

- 끝 -

APPENDIX 2. Questionnaire for the Students

※시험을 치르느라 수고하셨습니다. 여러분의 시험에 대한 의견을 듣고 싶습니다. 가능한 한 상세하게 답해주십시오. 설문지를 통해 수집된 여러분의 의견은 다음 정기고사 출제 시에 반영될 예정입니다. (Thank you for your effort in taking the exam. We'd like to listen to your opinions about the exam. Please answer as thoroughly as possible. Your opinions and suggestions on the exam may be applied to the next exam.)

Part 1: 시험 응시자 정보 (Test-taker information)

. 이름(name) _____ 성별(gender) _____

. 학번(student number) _____

. 수준별 반(proficiency level): 상(high)___ 중(mid)___ 하(low)___

. 이번 시험이 자신의 영어실력을 잘 반영하였다고 생각합니까? 아래에 동그라미 하십시오. (Do you think the exam accurately measured your English ability? Please circle your response below.)

< 전혀 아니다(strongly disagree) - 아니다(disagree) - 그저그렇다(neutral) - 그렇다(agree)
- 매우 그렇다(strongly agree) >

Part 2: 이제 각각의 문항에 대한 여러분의 의견을 듣고 싶습니다. 시험지를 펴놓고 각각의 문항에 대해서 아래 질문에 답하여 주십시오. (Now, we'd like to listen to your opinion about each item on the exam. Please take out your exam paper and answer each question presented below.)

문항 번호 (item no.)		전혀 아니다 (Strongly disagree)	아니다 (Disagree)	보통 이다 (Neutral)	그렇다 (Agree)	매우 그렇다 (Strongly agree)
1	문제는 어려웠다. (It was difficult.)					
	문제는 배운 내용을 잘 평가하였다. (It tested appropriately what was taught.)					
	문제의 형식이 적절하였다. (The form of the question was appropriate.)					
	문제점(Comments)					
2	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
3	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
4	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

문항 번호		전혀 아니다	아니다	보통 이다	그렇다	매우 그렇다
5	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
6	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
7	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
8	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
9	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

문항 번호		전혀 아니다	아니다	보통 이다	그렇다	매우 그렇다
10	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
11	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
12	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
13	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
14	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

문항 번호		전혀 아니다	아니다	보통 이다	그렇다	매우 그렇다
15	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
16	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
17	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
18	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
19	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

문항 번호		전혀 아니다	아니다	보통 이다	그렇다	매우 그렇다
20	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
21	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
22	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
23	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
서술 형 1	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

문항 번호		전혀 아니다	아니다	보통 이다	그렇다	매우 그렇다
서술 형 2	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
서술 형 3	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
서술 형 4	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					
서술 형 5	문제는 어려웠다.					
	문제는 배운 내용을 잘 평가하였다.					
	문제의 형식이 적절하였다.					
	(문제점)					

Part 3: 그 밖에 시험에 대한 일반적인 의견이 있다면 이곳에 써 주십시오. 성실한 답변에 감사드립니다. (Please write your opinions on the test in general. Thank you for your participation.)

APPENDIX 3-1. Questionnaire for the Teachers I

A. Testing Awareness and Practice

1. To what extent are you satisfied with your test writing ability? Do you think it has improved compared to the test that you first produced? What are your reasons?
2. What problems or concerns do you have when you construct test items?
3. Have you ever participated in classes / workshops / teacher trainings on testing? How was it helpful?
4. Have you ever received feedback from your students after the test? If not, do you think collecting students' feedback will be helpful to improve the test quality?
5. Do you think test writing ability is crucial in your career? Are you interested in developing your test writing ability?

B. The Midterm Exam

1. Consider the whole process of constructing the midterm exam. Did you have any difficulties or problems in the process of test writing?
2. How much are you satisfied with the items that you produced? How much are you satisfied with the final draft of the midterm exam?
3. Are there any items that are not very good in your view but included in the test? What items are good items from your perspective?

APPENDIX 3-2. Questionnaire for the Teachers II

1. Consider the whole process of constructing the final exam. Were there any changes compared to the midterm exam? Do you think those changes were affected by the first feedback session? What are they?
2. How much are you satisfied with the items that you produced? How much are you satisfied with the final draft of the final exam?
3. Do you think the quality of the final exam is better than that of the midterm exam? If so, in what perspective?

APPENDIX 3-3. Questionnaire for the Teachers III

1. What are the advantages of the feedback sessions that you have experienced?
In what way were they useful?
2. What was the most useful information that you obtained from the feedback sessions?
3. Do you have any comments regarding the method of the feedback sessions?
(ex. time, process, place, materials provided, etc)
4. Is there anything that you consider insufficient or unsatisfactory in the feedback sessions? Do you have any suggestions for the future feedback sessions?

국 문 초 록

본 연구는 특정학교에서 교사들이 집필한 시험문제의 질을 평가하고 시험문제의 질이 학생들의 시험에 대한 피드백에 의해 얼마나 향상될 수 있는가를 살펴보았다. 특히 본 연구는 1) 학생들이 교사들의 시험에 어떻게 반응하는지, 2) 학생들의 피드백이 교사들의 시험에 대한 인식과 시험 출제 과정에 어떤 영향을 미치는지, 3) 시험 문제의 질이 학생들의 피드백에 영향을 받아 실제로 더 나아지는지를 연구하였다. 약 760명의 고등학교 1학년 학생들이 중간고사와 기말고사 시험에 응시하였으며, 그 중 115명의 학생들이 시험 후 피드백 설문지를 작성하였다. 네 명의 고등학교 영어 교사들이 연구에 참여하였으며 그들은 시험 후에 두 번의 피드백 모임을 가지고 학생들의 피드백과 시험 결과를 살펴보았다. 본 연구의 주요 목적은 교사들의 시험에 대한 인식과 관습이 학생들의 피드백에 영향을 받아 변하는가를 살펴보는 것이었다.

전반적인 결과는 다음과 같다.

1. 학생들의 학교 시험 공정성(test fairness)에 대한 태도는 기말고사에서 나아졌음에도 불구하고 긍정적이지 않았다. 학생들이 인식한 각 문제에 대한 난이도, 변별도, 형식적합도 및 논평은 교사들이 그들의 시험을 개선할 수 있는 유용한 통찰력을 제공하였다. 학생들은 객관식 문제보다 서술형 문제에 더 많은 논평을 하였는데, 이는 서술형 문제가 더 많은 보완과 수정을 필요로 함을 암시한다.

2. 교사들의 시험에 대한 인식과 관습은 연구를 통해 많은 변화를 보여주었다. 학생들의 중간고사에 대한 피드백을 통해 교사들은 시험의 약점을 알

아내고 어떻게 고칠지를 의논하였다. 그들이 논의한 점을 기말고사 출제에 실행함으로써, 기말고사의 질이 실제로 향상되었다. 교사들은 피드백 회의가 그들의 시험이 어떠한지, 어떻게 고칠지, 나아가 그들의 노력이 어떤 결과를 가져올지를 깨닫게 해준다는 점에서 매우 긍정적인 태도를 보였다.

3. 시험 결과에 대한 분석은 기말고사의 객관식 문제가 질적으로 향상되었음을 보여주었다. 이것은 학생들의 피드백이 실제로 교사들이 그들의 시험을 개선하는 것을 도왔다는 것을 의미한다. 반면, 교사들의 서술형 문제 정답률을 높이고자 한 노력은 원하는 결과를 얻지 못하였다. 전반적으로, 교사들이 출제한 많은 문제들이 너무 쉽거나 너무 어려웠던 것으로 밝혀져 교사들의 시험 문제 개선에 대한 더 많은 노력이 필요함이 드러났다.

이러한 결과들은 학생들의 피드백이 교사들의 시험을 개선하는데 유용한 도구가 될 수 있음을 보여준다. 따라서 영어 교사들 사이에서 학생들의 피드백을 얻어, 이용하고자 하는 더 많은 노력이 필요하다. 교사들의 시험의 중요성 인식과 평가에 대한 꾸준한 맞춤형 교사 연수 및 훈련이 또한 이루어져야 하겠다.

주요어: 학생들의 피드백, 시험 개선, 교사들이 출제한 시험, 문항 분석,

교사들의 시험에 대한 인식과 관습

학 번: 2008-21572