General English Proficiency Test

Advanced Level Research Report

The Language Training and Testing Center

January, 2002
General English Proficiency Test

Advanced Level Research Report

Contents

I. GEPT Consultative, Advisory and Research Committees………………i

II. General English Proficiency Test Skill-Area Descriptions……………….ii

III. Preface……………………………………………………………………………….v

IV. Advanced Level Listening & Reading Test…………………………LR-1~LR-69

V. Advanced Level Writing Test………………………………………………W-1~W-30

VI. Advanced Level Speaking Test………………………………………..S-1~S-22

VII. Bibliography………………………………………………………………………..vii
GEPT Research
January, 2002

Consultative members of Ministry of Education (in alphabetical order):

Deputy Minister Fan, Sun-Lu
Deputy Minister Lu, Mu-lin
Deputy Minister Wu, Tieh-hsiung

General Director Chang, Kuo-Pao, Department of Higher Education
General Director Chen, De-hua, Department of Technological & Vocational Education
General Director Chou, Tsan-Der, Department of Social Education
General Director Hung, Ching-Hsiang, Department of Secondary Education
General Director Li, Chen-ching, Bureau of International Cultural & Educational Relations
General Director Liu, Yi-Chuan, Department of Elementary Education
Hsieh, Wen-ho, Chief, Division 1, Department of Social Education

Program Directors: Prof. Lin, Yaofu
Director Lu, Chen-lai

Advisory Committees (in alphabetical order):

Prof. Ahrens, Kathleen
Prof. Chang, Wu-chang Vincent
Prof. Chen, Mei-hua
Prof. Cheung, Hintat
Prof. Chou, Shu-hua
Prof. Huang, Can-sui
Prof. Huang, Pi-Twan
Prof. Huang, Tzyh-lai
Prof. Kao, Tien-en
Prof. King, Burt L.
Prof. Lee, Chingkwei Adrienne
Prof. Liao, Hsien-hao
Prof. Lin, Siieh-Hwa
Prof. Lin, Yaofu
Prof. Nash, Thomas
Prof. Shih, Yu-hwei
Prof. Su, I-wen Lily
Prof. Sze, Ding Kuo
Prof. Tsao, Feng-fu
Prof. Weir, Cyril J.
Vice Director Wu, R.W. Jessica
Prof. Yang, Tai-hsiung
Prof. Yin, Yuen-mei
Prof. Yu, Kuang-Hsiung

Research Committees (in alphabetical order):

Prof. Ahrens, Kathleen
Prof. Chang, Wu-chang Vincent
Prof. Cheung, Hintat
Prof. Chou, Shu-hua
Prof. Huang, Can-sui
Prof. Huang, Tzyh-lai
Prof. Sze, Ding Kuo
Vice Director Wu, R.W. Jessica
Prof. Weir, Cyril J.
Prof. Yang, Tai-hsiung
Prof. Yin, Yuen-mei
Prof. Yu, Kuang-Hsiung

Research Team:

R&D Department, The Language Training & Testing Center
Jessica R.W. Wu, Rachel Wu, Joyce Chin, Grace Kuo, Cecilia Liao, Echo Wei
General English Proficiency Test

General Level Descriptions

<table>
<thead>
<tr>
<th>Level</th>
<th>General Level Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>An examinee who passes this level has basic ability in English and can understand and use rudimentary language needed in daily life. His/her English ability is roughly equivalent to that of a junior high school graduate in Taiwan.</td>
</tr>
<tr>
<td>Intermediate</td>
<td>An examinee who passes this level can use basic English to communicate about topics in daily life. His/her English ability is roughly equivalent to that of a high school graduate in Taiwan.</td>
</tr>
<tr>
<td>High-Intermediate</td>
<td>An examinee who passes this level has a generally effective command of English; he/she is able to handle a broader range of topics, and although he/she makes mistakes, these do not significantly hinder his/her ability to communicate. His/her English ability is roughly equivalent to that of a university graduate in Taiwan whose major was not English.</td>
</tr>
<tr>
<td>Advanced</td>
<td>An examinee who passes this level has English ability which enable him/her to communicate fluently, with only occasional errors related to language accuracy and appropriateness, and to handle academic or professional requirements and situations.</td>
</tr>
<tr>
<td>Superior</td>
<td>An examinee who passes this level has English ability approaching those of a native English speaker who has received higher education. He/she is able to communicate effectively in all kinds of situations.</td>
</tr>
</tbody>
</table>

Skill-Area Descriptions

Elementary

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>An examinee who passes this level can understand simple conversation related to daily life on such topics as prices, time, and places.</td>
</tr>
<tr>
<td>Reading</td>
<td>An examinee who passes this level can understand simple written English related to daily life. He/she can read street signs, traffic signs, shop signs, simple menus, schedules, and greeting cards.</td>
</tr>
<tr>
<td>Writing</td>
<td>An examinee who passes this level can write simple sentences and paragraphs, such as those used in postcards, memos, and greeting cards. He/she can fill out forms and use simple written English to describe or explain topics related to daily life.</td>
</tr>
<tr>
<td>Speaking</td>
<td>An examinee who passes this level can read aloud simple passages and give a simple self-introduction. He/she can engage in simple dialogue in situations with which he/she is familiar, including greetings, shopping, and asking for directions.</td>
</tr>
</tbody>
</table>
**Intermediate**

<table>
<thead>
<tr>
<th>Language</th>
<th>Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listening</strong></td>
<td>An examinee who passes this level can understand general conversation in daily life situations and grasp the general meaning of public announcements, weather forecasts, and advertisements. At work, he/she can understand simple product introductions and operating instructions. He/she can catch the general meaning of native English speakers’ conversations and inquiries.</td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td>An examinee who passes this level can read short essays, short stories, personal letters, advertisements, leaflets, brochures, and instruction manuals. At work, he/she can read job-related information, company notices and operation manuals, as well as routine documents, faxes, telegrams and e-mail messages.</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td>An examinee who passes this level can write simple messages and narratives. He/she can write about things he/she has learned and use simple English to write about his/her own experiences or about topics with which he/she is familiar.</td>
</tr>
<tr>
<td><strong>Speaking</strong></td>
<td>An examinee who passes this level can use simple English to discuss or describe general daily life topics, introduce his/her daily life, work, family, and experiences, and state his/her outlook on general topics. At work, he/she can ask and answer basic questions and can carry on basic conversations with native English speakers in social settings.</td>
</tr>
</tbody>
</table>

**High-Intermediate**

<table>
<thead>
<tr>
<th>Language</th>
<th>Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listening</strong></td>
<td>An examinee who passes this level can understand conversations in social settings and grasp the general meaning of lectures, news reports, and TV/radio programs. At work, he/she can understand brief reports, discussions, product introductions, and operating instructions.</td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td>An examinee who passes this level can read written messages, instruction manuals, newspapers, and magazines. At work, he/she can read general documents, abstracts, meeting minutes, and reports.</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td>An examinee who passes this level can write general work-related reports and messages. In addition to topics related to daily life, he/she can write about current events and more complex or abstract subjects.</td>
</tr>
<tr>
<td><strong>Speaking</strong></td>
<td>An examinee who passes this level can fluently express his/her opinion about issues he/she is interested in. At work, he/she can socialize with native English speakers; explain the contents of a task or project; conduct business transactions and participate in discussions; give brief reports or express his/her opinions in a meeting.</td>
</tr>
</tbody>
</table>
### Advanced

| **Listening** | An examinee who passes this level can understand conversations on all sorts of topics as well as debates, lectures, news reports, and TV/radio programs. At work, when attending meetings or negotiations, he/she can understand reports and discussions. |
| **Reading** | An examinee who passes this level can understand all sorts of written English from a wide variety of sources, including magazine and newspaper articles, literature, professional periodicals, and academic publications. |
| **Writing** | An examinee who passes this level can use English appropriately in writing several text types, such as reports, essays, news items or summaries of general/professional topics. He/she can express his/her opinions on different topics and discuss them in depth. |
| **Speaking** | An examinee who passes this level can participate in discussions on, and fluently express his/her opinions about all sorts of issues. He/she can give reports or express his/her opinions in general meetings or professional seminars. |

### Superior

| **Listening** | An examinee who passes this level can understand spoken English on a wide variety of topics, from a wide variety of sources. |
| **Reading** | An examinee who passes this level can understand all sorts of written English on a wide variety of topics, from a wide variety of sources. |
| **Writing** | An examinee who passes this level can write many different types of documents, including proposals and reports, professional or academic abstracts, theses, news reports and editorials on current events. He/she can completely and effectively elaborate on, and carry out in-depth investigations into all sorts of subjects. |
| **Speaking** | An examinee who passes this level can express himself/herself in correct and fluent English in all kinds of situations. He/she can make appropriate use of cultural knowledge and idiomatic expressions. |
The Language Training and Testing Center (LTTC), under the support of the Ministry of Education (MOE), started to develop the General English Proficiency Test (GEPT) in 1999 to support MOE's lifelong learning policies. The development of all the five levels of the GEPT was scheduled to be completed by March 2002. The GEPT Intermediate Level Test was first launched in early 2000 and received positive feedback from the public. The GEPT High-Intermediate Level Test and the GEPT Elementary Level Test were launched in 2001. The number of examinees from a variety of age groups taking the GEPT has steadily increased year by year, and over 170,000 examinees have taken the GEPT in the year of 2002. GEPT test results are widely accepted in the public and private sectors as well as by schools at all levels, which indicates that the GEPT is highly recognized nationwide and has had a great impact on English teaching and learning in Taiwan. Recently, we have been getting a lot of inquiries about the launch of the GEPT Advanced and Superior Level Tests.

After years of effort, the GEPT research team has finally finished the development of the GEPT Advanced Level Test. From June 2000 to December 2001, the research team was engaged in designing test specifications and planning test content on the basis of the first and second stage GEPT research reports. To broaden our international vision and differentiate the Advanced Level Test from the other levels of the GEPT, Prof. Cyril James Weir, then Director of the Centre for Research in Testing and Evaluation, University of Surrey Roehampton, UK, was invited to join our GEPT research committee from January 2001 onwards. We'd like to express our thanks to Prof. Weir and the other research committee members for their guidance and supervision, which led to the successful completion of GEPT Advanced Level Test development.

The item types in the advanced listening, speaking, reading, and writing tests are different from those in the other GEPT level tests. In addition to conventional multiple-choice items, short answer questions and notes-completion items are used to assess advanced linguistic ability in a more diverse and integrated manner and add to content authenticity. Traditionally, reading assessment was viewed as one single construct, but this did not reflect reality since individuals adopt various reading strategies, depending on their purpose for reading. Therefore, the Advanced reading test includes careful reading (which tests thorough and full comprehension of articles), and skimming & scanning (which requires different reading strategies and tests whether candidates can quickly grasp the gist and main ideas of articles and locate specific information). The advanced writing test is designed to assess summary writing and general writing abilities in an integrated manner. The advanced speaking test has a
To examine the appropriateness of the advanced level test items, several small and large-scale pretests were conducted between April and October 2001. There were approximately 400 pretest examinees, including university seniors majoring in English (who were the target candidates of the GEPT Advanced Level Test) and individuals who had taken the GEPT High-Intermediate Level Test. Here we'd like to thank all of the schools, teachers, and students for their help and participation in our pretests. The pretest results showed that the design of the Advanced Level Test was generally appropriate. The Test is scheduled to be formally launched in March 2002.

This pretest report is divided into four sections, including listening, reading, speaking, and writing sections, each of which includes item design, statistical analysis of pretest performance, results of questionnaire surveys, analysis of items, conclusions and suggestions. It is hoped that the report can increase the public's understanding of the GEPT Advanced Level Test. Any advice or suggestions are welcome.
General English Proficiency Test

Advanced Level Listening & Reading Test

Contents

I. Pretest Development Process......................................................... LR-2
II. Descriptive Analysis of the Pretest Results................................. LR-8
III. Analysis of Sample Listening Pretest Items............................... LR-17
IV. Analysis of Sample Reading Pretest Items................................ LR-22
V. Conclusion and Suggestions...................................................... LR-27
VI. Appendices............................................................................. LR-29

Appendix 1: The GEPT Advanced Level Listening Pretest form............. LR-29
Appendix 2: The GEPT Advanced Level Reading Pretest form............... LR-47
Appendix 3: The GEPT Advanced Level Listening and Reading answer keys… LR-65
Appendix 4: Results of the student questionnaire................................ LR-67
I. Pretest Development Process

Based on results of the first and second stage research conducted in 1997 by the GEPT Research Committee, the can-do descriptors of the GEPT Advanced Listening and Reading levels were in the following:

Listening: Examinees who pass the GEPT Advanced Level Listening Test should be able to comprehend talks, debates, lectures, reports and programs on a variety of topics; while attending business meetings or negotiations in the workplace, they can comprehend what is being reported or discussed.

Reading: Examinees who pass the GEPT Advanced Level Reading Test should be able to read texts in different forms and on a variety of topics, including news and magazine articles, literary works, professional journals, academic writing and literature.

The research team designed the GEPT Advanced Level Listening and Reading Test specifications, compiled an item writing manual, and developed pretest items, with reference to previous studies at home or abroad, interviews with university lecturers at English language-related departments, and content analysis of widely-used English teaching materials. The research team created a prototype form in early 2001 and conducted the first pretest in May 2001. A revision of the prototype form was made on the basis of the pretest results analysis and the comments of our research committee members before the second pretest was conducted in October 2001. The present report focused on the analysis of the second pretest, serving as a model reference for future test compilation.

1. Listening test item types

The listening test consisted of forty items; the full score was 120 points, and each item was equal weighting. The duration of the test was approximately forty-five minutes. The items in Part I: Short Conversations & Talks were multiple-choice questions with four choices. Apart from a few multiple-choice questions, most items in Part II: Long Conversations and Part III: Long Talks were short answer questions or notes-completion. Table 1 shows the item type and the test focus\(^1\) of each part. For the listening pretest items, please refer to Appendix 1.

---

\(^1\) The test focus of this test was created with reference to Weir's listening comprehension framework (1993) and the empirical listening test studies of Buck (1991), Shohamy & Inbar (1991), and Tsui & Fullilove (1998).
Table 1  *The item types of the listening test*

<table>
<thead>
<tr>
<th>Part</th>
<th>Item type</th>
<th>Item format</th>
<th>No. of items</th>
<th>Duration</th>
<th>Test focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>Short Conversations &amp; Talks</td>
<td>Multiple-choice questions</td>
<td>15</td>
<td></td>
<td>To assess examinees' ability to understand extended discourse; to assess whether examinees can:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1) understand the topic and main ideas of conversations and talks;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) understand contextual features, such as the purpose of a conversation, the location, and the speakers' relationship;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) understand important facts and supporting details;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) make inferences from the information heard, such as the implications, attitudes, or intentions of the speakers in a conversation.</td>
</tr>
<tr>
<td>Part II</td>
<td>Long Conversations</td>
<td>1) Multiple-choice questions</td>
<td></td>
<td>45 mins</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Short answer questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Notes-completion</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part III</td>
<td>Long Talks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The item types were all arranged in order of ascending difficulty level. Part I began with short conversations and talks with multiple-choice questions so as to ease examinees' anxiety. The content (conversations or talks) was recorded, spoken in turn by a male and a female at a normal speed\(^2\) of approximately 190 wpm with an American accent and was played JUST ONE TIME. Examinees were allowed to make notes in their test booklet while listening to prevent their short-term memory loss from affecting their listening comprehension. The test format of each part was in the following:

**Part I: Short Conversations & Talks:** This part of the test consisted of 15 short conversations and talks, and the length of each was forty to seventy words. At the end of each conversation and talk, there would be a question about what was said. Examinees should read the four choices in the test booklet and choose the best answer to the question they have heard. The amount of time provided for responding to each question was approximately ten seconds.

**Parts II & III: Long Conversations & Talks:** These two parts of the test consisted of two long conversations and two long talks (such as lectures, presentations, news reports, etc.). The length of each conversation or talk was 400 to 500 words, followed by five to seven related short answer questions, notes-completion questions or multiple-choice questions. There were a total of twenty-five questions. Before listening to each conversation or talk, examinees would have 30 seconds to read the questions which were printed in the test booklet. After hearing each conversation or talk, examinees would have three to four minutes (depending on the number of the questions) to answer the questions.

---

\(^2\) Pimsleur, et al. (1977) divide the speed of English speaking into several categories: Fast: above 220 wpm, Moderately fast: 190-220 wpm, Average: 160-190 wpm, Moderately slow: 130-160 wpm, and Slow: below 130 wpm. Wpm stands for "words per minute."
2. Listening test contents and item writing principles

The selection of advanced listening contents were based on the can-do descriptors mentioned above. The item difficulty should be aligned with what English majors have learned during the four years of the undergraduate program. The topics and situational contexts, at both local and global level, should avoid containing too professional or too unusual topics. The mean facility value of the listening test items was set at 0.60. In addition, the studies of Shohamy & Inbar (1991), Berne (1992), and Hui-chun Teng (1998) showed that situational contexts, discourse/text types, and orality of listening tests have a significant impact on examinees' performance. Therefore, the research team had put much emphasis on content diversity, balance, and authenticity while designing listening test items. Table 2 shows the planning of the test contents:

Table 2  The categories of listening test contents

<table>
<thead>
<tr>
<th>Part</th>
<th>Item type</th>
<th>Content category</th>
<th>Situational context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>Short Conversations &amp; Talks</td>
<td>Everyday conversations, discussions and information transactions in the workplace, short narrations/descriptions, lectures, commentaries, news reports, advertisements, etc.</td>
<td>1) Everyday life: social conversations, family, residence, transportation, entertainment, medical health, current affairs, and popular topics.</td>
</tr>
<tr>
<td>Part II</td>
<td>Long Conversations</td>
<td>Interviews, long discussions, broadcast programs, etc.</td>
<td>2) School life: curriculum, learning situations, extra-curricular activities, academic themes (humanities, arts, natural sciences, social sciences, etc.)</td>
</tr>
<tr>
<td>Part III</td>
<td>Long Talks</td>
<td>Lectures, presentations, news reports, broadcast programs, etc.</td>
<td>3) Workplace: job seeking, work instructions, business dealings, interpersonal interactions at work.</td>
</tr>
</tbody>
</table>

The focus of the pretest development was not only on selection of test contents but also design of comprehension questions. According to the test focus shown in Table 1, the comprehension questions could be generally classified into three categories: 1) global comprehension: to assess whether examinees could grasp the topic, purpose, main ideas and framework of conversations and talks, and understand contextual features (such as occasion, location, and speakers' relationship); 2) local comprehension: to assess whether examinees could grasp important facts and supporting details of conversations and talks; 3) inference-making: to assess whether examinees could make inferences from the information heard, such as the implications, attitudes, or intentions of the speakers in a conversation.

In addition, as the only one of its kind among the GEPT listening tests, the Advanced Listening Test adopted short answer questions, notes-completion questions and heading matching questions in its Part II and Part III for the purpose of assessing listening comprehension ability in a diverse and integrated way and making item writing reflect realities more properly. The test specifications for non-multiple-choice items were carefully designed by the research team, with reference to relevant theory and empirical studies and analysis of similar tests developed by international testing institutes. In general, item writing
for short answer questions and notes-completion questions should pay special attention to clarity of lexis to avoid examinees' confusion about the meaning of questions or to prevent marking difficulties arising from multiple answers to one question, which would be discussed further in Item Analysis.

3. Reading test item types

The reading test consisted of forty items; the full score was 120 points, with three points allocated to each of the items. The duration of the test was approximately seventy minutes, with different amounts of time provided for responding to each of the two parts. Examinees were allowed fifty minutes to answer the short answer questions, summary-completion questions and multiple-choice questions in Part I: Careful Reading, and another twenty minutes to answer heading matching questions and 3-way multiple matching questions in Part II: Skimming and Scanning. Table 3 shows the item type and the test focus of each part. For the reading pretest items, please refer to Appendix 2.

Table 3  The item types of the reading test

<table>
<thead>
<tr>
<th>Part</th>
<th>Item type</th>
<th>Item format</th>
<th>No. of items</th>
<th>Duration (mins)</th>
<th>Text length</th>
<th>Reading rate</th>
<th>Test focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>Careful Reading</td>
<td>❖ Short answer questions</td>
<td>20</td>
<td>50</td>
<td>About 3000 words in total</td>
<td>80-100 wpm</td>
<td>To assess whether examinees can grasp gist, main ideas, and important details of a variety of articles, understand contextual features and structure, and make inference about authors' implications, viewpoints, and attitude.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>❖ Summary Completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice questions (Part 2A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part II</td>
<td>Skimming &amp; Scanning</td>
<td>Skimming</td>
<td>20</td>
<td>20</td>
<td>About 2500 words in total</td>
<td>150-200 wpm</td>
<td>To assess whether examinees can browse articles by applying reading strategies so that they can quickly grasp gist and main ideas of articles and look up specific information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heading matching (Part 2B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scanning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3-way multiple matching (Part 2C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The reading test mainly featured the separation of careful reading and skimming & scanning. There was difference between the two individual sections in the test focus, required reading rate, item format, and the amount of time provided for responding to questions. The special item design was created with reference to Urquhart & Weir's reading and
comprehension framework (1998) and the empirical study of Weir et al. (2000). These researchers pointed out that traditional reading tests overemphasized careful reading, placing reading ability in a unitary dimension, and maintained that reading comprehension involved interaction between reading techniques and strategies of different levels, and that how readers used their reading techniques and strategies often varied with their reading purposes. Therefore, in teaching and assessing reading ability, not only careful reading but also skimming and scanning techniques should be included. As for the purposes and patterns of reading, careful reading aims to achieve thorough and deep understanding, and readers are often required to read every word and sentence. Skimming aims to grasp main ideas of texts, and scanning aims to search for specific information quickly; therefore, readers often skip unimportant parts of texts and read only specific parts.

In addition, as for reading rates, the careful reading in this pretest required a reading rate of 80-100 wpm, and skimming and scanning of 150-200 wpm. The required rates were determined with reference to the study of Weir et al. (2000), the requirements of China's Test for English Majors-Grades 4 & 8, and the performance of our first-stage pretest examinees. The test format of the Advanced Reading pretest is detailed below:

Part I (Careful Reading): This part consisted of four texts with twenty questions in total. Each text contained 600 to 900 words, followed by four to seven related short answer questions, summary-completion questions and a few multiple-choice questions.

Part II (Skimming & Scanning): This part consisted of three sections with a total of twenty questions.

Part 2A (Main ideas of short passages) : Part 2A consisted of four short passages with six questions. Each passage contained 150 to 300 words, preceded by one to two related questions. Examinees had to first read the questions and then read the passages quickly and selectively to find the answers.

Part 2B (Main ideas of paragraphs) : Part 2B consisted of one passage with six questions altogether. The passage contained 600 to 800 words, divided into several paragraphs. Each paragraph had a heading that expressed the main idea of that paragraph. Examinees had to read the paragraphs quickly and choose one of the headings from the list for each paragraph that lacked a heading.

Part 2C (Scanning for specific information) : Part 2C consisted of three passages on the same topic with eight questions altogether. Each passage contained approximately 300 words. Examinees had to first read the questions and then read the passages quickly and selectively to find the answers.
4. Source texts and item writing principles for reading tests

Source texts chosen for advanced level reading tests should meet the required reading ability descriptor, seek a balance of topics and text types, and achieve a mean F.V. of 0.60, just as advanced level listening tests do. In addition, the difference between Part I (careful reading) and Part II (skimming and scanning) in the test focus, reading rate, and item type resulted in differences between the two parts in choice of topics, length of texts, text types, and information structures. The range of source texts and topics for each part is detailed in Table 4:

Table 4  *The range of source texts and topics in advanced level reading tests*

<table>
<thead>
<tr>
<th>Part</th>
<th>Item type</th>
<th>Literary form</th>
<th>Topic</th>
<th>Source Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>Careful Reading</td>
<td>descriptive, narrative, expository, and argumentative articles</td>
<td>Three categories</td>
<td>Newspapers, magazines, professional journals, general books, academic works, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1) Humanities: arts, literature, history, education, commerce, and psychology, etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) Science and technology: natural sciences, computer technology, etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3) Everyday life: transportation, entertainment, medical health, environmental protection, current affairs, popular topics, etc.</td>
<td></td>
</tr>
<tr>
<td>Part II</td>
<td>Skimming &amp; Scanning</td>
<td>narrative, expository, and argumentative articles</td>
<td>Three categories</td>
<td>Newspapers, magazines, general books, flyers, manuals, advertisements, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>narrative and expository articles</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the test focus shown in Table 3, Part I (careful reading) generally focused on global comprehension, local comprehension, and inference making. The test focus of Part I was quite similar to that of the listening test. However, items that tested global comprehension had to effectively assess examinees' ability to consider key information as a whole, while items that tested local comprehension and inference making had to effectively assess examinees' ability to properly interpret texts. In addition, the item types of this reading test included short answer questions, summary-completion, and multiple-choice questions; the design of item types and the principles of item writing for the reading test were similar to those for the listening test. Therefore, special attention had to be paid to clarity of items to avoid multiple answers.

As for Part II (skimming and scanning), the first two item types, multiple-choice questions and heading matching, aim to assess whether examinees can quickly grasp main ideas of texts and gist of paragraphs by using skimming techniques. The third item type, 3-way multiple matching, aims to assess whether examinees can use scanning skills properly and focus on factual information provided in texts. In addition, questions were printed above the related texts. In this way, examinees could set their reading purposes before reading the texts, making this item type more authentic.
5. Scoring methods

Both the listening test and the reading test consisted of multiple-choice items and non-multiple-choice items. Multiple-choice items were computer scored, while non-multiple-choice items (short answer questions, and summary-completion questions) were human scored. Scoring would be determined by the key words used in answers. Scores were not deducted for errors in spelling and grammar that did not prevent understanding answers. On the contrary, errors in spelling and grammar that greatly impeded markers' understanding would result in zero. The steps of scoring non-multiple-choice items in this pretest are explained below:

Step 1: making of answer keys
The answer keys of this listening and reading tests were created on the basis of the answers provided by the item writers, with reference to the answers given by the two native English speakers who took the pretest. The key information of the answer keys was italicized.

Step 2: trial marking
Forty samples were selected from 234 test papers after the pretest and were trial marked according to the answer keys. During the trial marking, synonymous answers (for example, 'cost of insurance' was synonymous with 'insurance cost') were collected and recorded, and those that were dubious or seemingly right were under discussion until consenses were reached.

Step 3: making of a marking scheme
The research team expanded the original answer keys into a marking scheme, in which synonymous answers and the ways of dealing with dubious answers were included.

Step 4: formal marking
The research team marked the remaining test papers according to the marking scheme. New dubious answers were still discussed during formal marking.

II. Descriptive Analysis of the Pretest Results

1. Summary of statistical results

A total of 234 persons took this pretest. The statistical results of the listening performance and the reading performance are shown in Tables 5 and 6. (The results were calculated on a raw score basis, with three points allocated to each item. The full score was 120 points.)
Table 5  Statistical results of the listening performance

<table>
<thead>
<tr>
<th></th>
<th>Part I (Short Conversations &amp; Talks)</th>
<th>Part II (Long Conversations)</th>
<th>Part III (Long Talks)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of items</td>
<td>15</td>
<td>12</td>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>No. of correct items</td>
<td>7.95</td>
<td>6.94</td>
<td>7.19</td>
<td>21.85</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.95</td>
<td>2.52</td>
<td>3.27</td>
<td>7.65</td>
</tr>
<tr>
<td>Mean facility value</td>
<td>0.53</td>
<td>0.58</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>Maximum correct items</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>Minimum correct items</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Reliability (Alpha)</td>
<td>0.67</td>
<td>0.69</td>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>Standard error</td>
<td>1.71</td>
<td>1.40</td>
<td>1.50</td>
<td>2.65</td>
</tr>
</tbody>
</table>

N = 234

Table 6  Statistical results of the reading performance

<table>
<thead>
<tr>
<th></th>
<th>Part I (Careful Reading)</th>
<th>Part II (Skimming &amp; Scanning)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of items</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>No. of correct items</td>
<td>12.08</td>
<td>10.77</td>
<td>22.61</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.65</td>
<td>3.88</td>
<td>6.77</td>
</tr>
<tr>
<td>Mean facility value</td>
<td>0.60</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td>Maximum correct items</td>
<td>19</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>Minimum correct items</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Median</td>
<td>13</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>Reliability (Alpha)</td>
<td>0.73</td>
<td>0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>Standard error</td>
<td>1.89</td>
<td>1.98</td>
<td>2.79</td>
</tr>
</tbody>
</table>

N = 234

1) Analysis of item difficulty: The analysis was based on the \( p \) value, namely the facility value.

The overall mean F.V. of the listening test was 0.55. Part I (Short Conversations & Talks) turned out to be more difficult and therefore achieved a lower mean F.V. (0.53) than Part II (0.58) and Part III (0.55), which did not meet the research team's expectation that items should be arranged in order of ascending difficulty level. Therefore, item writing for Part I needed to be improved, and this will be further discussed in item analysis.

The overall mean F.V. of the reading test was 0.57. Part II (Skimming & Scanning) was more difficult than Part I (Careful Reading) and achieved a lower mean F.V. (0.54) than Part I (0.60). According to the survey results, most of the examinees had never taken Skimming & Scanning tests before but were familiar with Careful Reading. As a result, the reading performance was considered acceptable.
2) Analysis of discrimination indices (D.I.): The D.I. analysis was based on point-biserial correlation.

In terms of the listening test, only one item achieved a D.I. lower than 0.20. The D.I.s of the other thirty-nine items were all above 0.20. As for the reading test, 85% (34 items) of the items achieved a good D.I., while six items achieved a D.I. lower than 0.20.

3) Analysis of reliability:

Reliability is the extent to which test results are consistent and stable. The Alpha value of the internal consistency of the listening and reading tests (a total of eighty items) was 0.91. The reliability index of the listening test (a total of forty items) was 0.88, and that of the reading test (a total of forty items) was 0.83. All of the reliability indices were about 0.80. The overall reliability of the reading test was slightly lower than that of the listening test, which might be due to the difference in test focuses between Careful Reading and Skimming & Scanning. The reliability indices of Careful Reading (a total of twenty items) and Skimming & Scanning (a total of twenty items) were 0.73 and 0.74, respectively.

The standard error of measurement (SEM) of the listening test was 2.65 items (7.95 points), meaning that when examinees re-took the same test, the probability that their true scores would range between 7.95 points higher and lower than the actual scores was 68%. The overall SEM of the reading test was 2.769 items (8.37 points), which was quite close to that of the listening test.

2. Score distributions and estimates of passing rates

As planned for the GEPT Advanced Level tests, the cut score of the two tests of the first stage (listening and reading tests) should be 80 out of a total score of 120, respectively. Each item received three points, and examinees had to answer more than twenty-seven items correctly (81 scores) in both tests to pass the test. Table 7 shows the score distributions of the listening and reading tests. The passing rate for the listening test was 35%, while that for the reading test was also 35%. Table 8 shows that a total of fifty-eight examinees, accounting for 25% of the examinees, passed both tests.
<table>
<thead>
<tr>
<th>No. of correct items</th>
<th>Listening Test</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Reading Test</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examinees (N)</td>
<td>Cumulative total</td>
<td>Cumulative %</td>
<td>Examinees (N)</td>
<td>Cumulative total</td>
<td>Cumulative %</td>
<td>Examinees (N)</td>
<td>Cumulative total</td>
<td>Cumulative %</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>7</td>
<td>16</td>
<td>7</td>
<td>4</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>6</td>
<td>22</td>
<td>9</td>
<td>5</td>
<td>17</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>6</td>
<td>28</td>
<td>12</td>
<td>11</td>
<td>28</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>9</td>
<td>37</td>
<td>16</td>
<td>11</td>
<td>39</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>47</td>
<td>20</td>
<td>11</td>
<td>50</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>14</td>
<td>61</td>
<td>26</td>
<td>11</td>
<td>61</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>11</td>
<td>72</td>
<td>31</td>
<td>7</td>
<td>68</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>27 (67.5%)</td>
<td><strong>10</strong></td>
<td><strong>82</strong></td>
<td><strong>35%</strong></td>
<td><strong>13</strong></td>
<td><strong>81</strong></td>
<td><strong>35%</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>89</td>
<td>38</td>
<td>12</td>
<td>93</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>9</td>
<td>98</td>
<td>42</td>
<td>10</td>
<td>103</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>10</td>
<td>108</td>
<td>46</td>
<td>13</td>
<td>116</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>10</td>
<td>118</td>
<td>50</td>
<td>5</td>
<td>121</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>124</td>
<td>53</td>
<td>12</td>
<td>133</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>133</td>
<td>57</td>
<td>13</td>
<td>146</td>
<td>62</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>142</td>
<td>61</td>
<td>11</td>
<td>157</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>12</td>
<td>154</td>
<td>66</td>
<td>13</td>
<td>170</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>14</td>
<td>168</td>
<td>72</td>
<td>11</td>
<td>181</td>
<td>77</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>178</td>
<td>76</td>
<td>10</td>
<td>191</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>189</td>
<td>81</td>
<td>8</td>
<td>199</td>
<td>85</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>194</td>
<td>83</td>
<td>4</td>
<td>203</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>201</td>
<td>86</td>
<td>8</td>
<td>211</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>207</td>
<td>88</td>
<td>10</td>
<td>221</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>211</td>
<td>90</td>
<td>6</td>
<td>227</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>220</td>
<td>94</td>
<td>2</td>
<td>229</td>
<td>98</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>222</td>
<td>95</td>
<td>0</td>
<td>229</td>
<td>98</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>224</td>
<td>96</td>
<td>3</td>
<td>232</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>228</td>
<td>97</td>
<td>2</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>231</td>
<td>99</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>232</td>
<td>99</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>233</td>
<td>100</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>233</td>
<td>100</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>234</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
3. Analysis of test performance

The examinees who took the listening and reading tests consisted of two categories: 1) university seniors who majored in English and 2) those who passed the second GEPT High-Intermediate first-stage test conducted in 2001 (ninety-five of them passed the HI second-stage test and the remaining forty-eight failed.) Table 8 shows the statistical results of the performance of these examinees.

Table 8  Statistical results of the listening and reading performance

<table>
<thead>
<tr>
<th>Category of examinees</th>
<th>Examinees (N)</th>
<th>Mean scores</th>
<th>Those who passed both tests (≥27 items)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Listening</td>
<td>Reading</td>
</tr>
<tr>
<td>Category 1 (total)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Seniors majoring in English)</td>
<td>91</td>
<td>67.82</td>
<td>70.21</td>
</tr>
<tr>
<td>Those who passed the HI second-stage test</td>
<td>95</td>
<td>74.31</td>
<td>74.69</td>
</tr>
<tr>
<td>Those who did not pass the HI second-stage test</td>
<td>48</td>
<td>48.54</td>
<td>54.42</td>
</tr>
<tr>
<td>Category 2 (total)</td>
<td>143</td>
<td>65.66</td>
<td>67.89</td>
</tr>
<tr>
<td>(Those who passed the GEPT High-Intermediate first-stage test)</td>
<td></td>
<td>(98.27)</td>
<td>(94.81)</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>66.50</td>
<td>68.79</td>
</tr>
</tbody>
</table>

Note: the figures in parentheses refer to mean scores of the HI tests.

The first-category examinees (a total of ninety-one examinees) obtained higher mean scores in the two tests than the second-category examinees (a total of 143 examinees), but lower mean scores than those who passed the HI second-stage test (a total of ninety-five examinees). Of the first-category examinees, 30% passed both listening and reading tests, while 33% of those who passed the HI second-stage test passed both tests.

In addition, to examine the difference in item difficulty between the Advanced test and the HI test, the performance of the 143 examinees on the HI first-stage test and that on the Advanced first-stage test were compared, and the results are listed below:

1) The mean listening score of the 143 examinees was 98.27 in the HI test and 65.66 in the Advanced test, with a difference of 32.61 points. The mean reading score was 94.81 in the HI test and 67.89 in the Advanced test, with a difference of 26.92 points.
2) With a cut score of 80 ( ≥27 correct items) for both listening and reading tests, thirty-one of the 143 examinees passed the first stage of the Advanced test; the passing rate was 33%. All of the thirty-one examinees had passed the HI second-stage test. None of the examinees who failed the HI second-stage test passed the Advanced first-stage test.
The above results showed that the first stage of the Advanced level tests discriminated well. There was a difference of 25-30 points in the difference of the mean scores of the listening and reading tests between the HI test and the Advanced test.

Also, according to the survey results, forty-three examinees had taken paper-based TOEFL tests in the preceding two years. Table 9 shows the pretest performance of these examinees divided into three categories. Generally, their performance in the listening and reading tests was positively correlated with their TOEFL results, indicating this pretest could adequately discriminate examinees of different English levels. Those who obtained a TOEFL score of 600 or above achieved a higher passing rate in the first stage of the Advanced test, accounting for 60% of the total examinees who passed the first stage of the Advanced test. However, the sample (five examinees) was very small. Larger samples need to be collected in the future to verify the above analysis.

Table 9  
**Performance of TOEFL examinees in the Advanced preliminary test**

<table>
<thead>
<tr>
<th>Category of examinees</th>
<th>Examinees (N)</th>
<th>Mean score</th>
<th>Those who passed both tests (≥27 items)</th>
<th>% (of the same-category examinees)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Listening</td>
<td>Reading</td>
<td>Examinees (N)</td>
</tr>
<tr>
<td>TOEFL PBT³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;550</td>
<td>16</td>
<td>57.13</td>
<td>63.25</td>
<td>0</td>
</tr>
<tr>
<td>550-600</td>
<td>22</td>
<td>67.23</td>
<td>69.09</td>
<td>4</td>
</tr>
<tr>
<td>600-650</td>
<td>5</td>
<td>86.80</td>
<td>79.00</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>65.74</td>
<td>68.07</td>
<td>7</td>
</tr>
</tbody>
</table>

4. Correlation

Table 10 shows correlation coefficients within different parts of the listening and reading tests.

Table 10  
**Correlation coefficients between parts of the listening and reading tests**

<table>
<thead>
<tr>
<th></th>
<th>Listening Part 1</th>
<th>Listening Part 2</th>
<th>Listening Part 3</th>
<th>Reading Part 1</th>
<th>Reading Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Part 1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Part 2</td>
<td>0.58</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Part 3</td>
<td>0.63</td>
<td>0.72</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Part 1</td>
<td>0.47</td>
<td>0.60</td>
<td>0.58</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reading Part 2</td>
<td>0.28</td>
<td>0.38</td>
<td>0.34</td>
<td>0.61</td>
<td>1</td>
</tr>
</tbody>
</table>

N =234

The overall correlation between the listening test and the reading test was 0.66, with 234 examinees in total taking both tests. As for the correlations between the three parts of the

³ According to the survey results (Question 2), only fourteen pretest examinees had taken TOEFL CBT tests. The sample size was too small to be representative and, thus, omitted here. The score comparison between TOEFL PBT and CBT tests are listed below: PBT 500 = CBT 173, PBT 550 = CBT 213, PBT 600 = CBT 250, and PBT 650 = CBT 280.
listening test (convergent validity\(^4\)), the highest was between Part II (Long Conversations) and Part III (Long Talks), which was probably because these two parts had similar text length and used both short answer questions and notes-completion. The correlation between Part I (Careful Reading) and Part II (Skimming & Scanning) of the reading test was 0.61, indicating that the abilities tested were relevant to each other rather than identical.

As for the correlation coefficients within different parts of the listening and reading tests (discriminant validity\(^5\)), those between each part of the listening test and Part I (Careful Reading) of the reading test were generally higher (0.47-0.60) than those between each part of the listening test and Part II (Skimming & Scanning) of the reading test (0.28-0.38).

5. Results of questionnaire survey

At the end of the pretest, questionnaires were circulated to examinees. A total of 225 valid returns were received. The results of the questionnaire survey are summarized below. For the complete results, please refer to Appendix 3.

1) The listening test

Instructions for each part of this listening test: 72% of the examinees considered the instructions to be clear, 25% acceptable, and 2% unclear.

Speech rate: 61% of the examinees considered the speech rate to be just right, while 37% felt it was slightly fast, and 1% slightly slow.

Number of test items: 67% of the examinees felt there were the right number of test questions, 15% slightly too few, and 17% slightly too many.

Time for responding to the questions (Part I - ten seconds; Part II/III – three to four minutes): More than a half of the examinees considered the amount of time to be just right, while nearly 40% considered it to be slightly short.

Time for a preview of Part II/III (approximately thirty seconds): 62% of the examinees considered the time to be just right, 28% slightly short, and 10% slightly long.

Perceived difficulty of item types: Generally, the examinees considered the item types to be arranged in order of ascending difficulty level, which did not match the order of the F.V. results (Part I had the lowest F.V., and Part II had the highest F.V.). Table 11 shows the survey results.

\(^4\) According to Campbell & Fiske (1959), convergent validity is the degree to which test concepts that should be related theoretically are interrelated in reality.

\(^5\) According to Campbell & Fiske (1959), discriminant validity is the degree to which test concepts that should not be related theoretically are not interrelated in reality.
Table 11

<table>
<thead>
<tr>
<th></th>
<th>Easy</th>
<th>Just right</th>
<th>Slightly difficult</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>7%</td>
<td>47%</td>
<td>42%</td>
<td>4%</td>
</tr>
<tr>
<td>Part II</td>
<td>4%</td>
<td>45%</td>
<td>42%</td>
<td>8%</td>
</tr>
<tr>
<td>Part III</td>
<td>3%</td>
<td>38%</td>
<td>44%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Whether the test is a valid measure of listening proficiency: 43% of the examinees fully agreed that the listening test could adequately assess their listening ability, 49% barely agreed, and 9% disagreed. The survey results showed that the face validity of this test was acceptable.

Expectations of test performance: Table 12 shows that there was a difference between examinees' expectations and their test performance. For example, 41% of the examinees expected to obtain an F.V. below 0.60, but actually 54% of the examinees obtained an F.V. below 0.60, which indicated that the low scoring group tended to overestimate their performance.

Table 12

<table>
<thead>
<tr>
<th>Expected F.V.</th>
<th>Examinees (%)</th>
<th>Actual F.V.</th>
<th>Examinees (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above 80%</td>
<td>11%</td>
<td>Above 80%</td>
<td>12%</td>
</tr>
<tr>
<td>60–80%</td>
<td>48%</td>
<td>60–80%</td>
<td>34%</td>
</tr>
<tr>
<td>Below 60%</td>
<td>41%</td>
<td>Below 60%</td>
<td>54%</td>
</tr>
</tbody>
</table>

2) The reading test

The item type, Skimming & Scanning, was first used in this advanced-level reading test. Although the examinees had been provided with detailed instructions for the test method and notified of the test duration, many of them still read the texts word by word in Part II (Skimming & Scanning) and thus failed to complete all of the questions in time. The survey results also reflected this situation. Below is the summary of the survey results:

Instructions for each part of this reading test: 63% of the examinees considered the instructions to be clear, 29% acceptable, and 8% unclear.

Table 13-1  Number of test items

<table>
<thead>
<tr>
<th></th>
<th>Slightly too few</th>
<th>Just right</th>
<th>Slightly too many</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I (Careful Reading)</td>
<td>3%</td>
<td>58%</td>
<td>38%</td>
</tr>
<tr>
<td>Part II (Skimming &amp; Scanning)</td>
<td>1%</td>
<td>22%</td>
<td>75%</td>
</tr>
</tbody>
</table>
Table 13-2  *Time for responding to the questions*

<table>
<thead>
<tr>
<th></th>
<th>Slightly short</th>
<th>Just right</th>
<th>Slightly long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I (50 mins)</td>
<td>49%</td>
<td>45%</td>
<td>6%</td>
</tr>
<tr>
<td>Part II (20 mins)</td>
<td>80%</td>
<td>16%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 13-3  *Perceived difficulty of item types*

<table>
<thead>
<tr>
<th></th>
<th>Easy</th>
<th>Just right</th>
<th>Slightly difficult</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>4%</td>
<td>47%</td>
<td>40%</td>
<td>9%</td>
</tr>
<tr>
<td>Part II</td>
<td>3%</td>
<td>45%</td>
<td>39%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 13-4  *Whether the test is a valid measure of reading proficiency*

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>Barely agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>39%</td>
<td>49%</td>
<td>11%</td>
</tr>
<tr>
<td>Part II</td>
<td>27%</td>
<td>44%</td>
<td>27%</td>
</tr>
</tbody>
</table>

As shown in Tables 13-1, 13-2 and 13-4, most of the examinees considered there were too many questions to finish in the allotted time. Also, 27% of the examinees disagreed that this test could adequately assess their reading ability. However, it was noteworthy that although nearly 80% of the examinees expressed their failure to finish the questions in Part II (Skimming & Scanning), 45% of them still considered the difficulty of this part to be just right, which was quite close to the 'just right' percentage (47%) of Part I (Careful Reading). It indicated that the examinees did not consider Skimming & Scanning to be difficult but just could not get used to changes in reading strategies and the time limit.

Expectations of test performance: Table 14 shows that stronger examinees tended to be less optimistic about their reading performance. For example, only 15% of the examinees expected to correctly answer 80% of the items in Part I, while in fact 23% of the examinees made it. As for Part II, only 5% of the examinees expected to correctly answer 80% of the items, while 13% of the examinees made it.

Table 14

<table>
<thead>
<tr>
<th>Expected % of items answered correctly</th>
<th>Examinees %</th>
<th>Actual F.V.</th>
<th>Examinees %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part I</td>
<td>Part II</td>
<td>Part I</td>
</tr>
<tr>
<td>Above 80%</td>
<td>15%</td>
<td>5%</td>
<td>23%</td>
</tr>
<tr>
<td>60-80%</td>
<td>46%</td>
<td>35%</td>
<td>38%</td>
</tr>
<tr>
<td>Below 60%</td>
<td>37%</td>
<td>54%</td>
<td>39%</td>
</tr>
<tr>
<td>Unanswered</td>
<td>2%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>
III. Analysis of Sample Listening Pretest Items

Analysis of sample listening pretest items and examinees' performance are presented below. Also, suggestions about how to improve item writing are provided as reference for future test compilation. The quantitative data analysis included the $p$ value (facility value), discrimination index (D.I.), the F.V. of both the high and low scoring groups, and distractor analysis. As for the qualitative analysis, the research team not only reviewed theoretical frameworks and previous studies of listening comprehension but also collected opinions and comments from examinees, English lecturers at universities, and our committee members.

1. Part I (Short Conversations & Talks)

Table 15  Distribution of F.V.s of Part I in the listening test

<table>
<thead>
<tr>
<th>Sections/Items</th>
<th>Too difficult</th>
<th>Slightly too difficult</th>
<th>Moderately difficult</th>
<th>Moderately easy</th>
<th>Slightly too easy</th>
<th>Too easy</th>
<th>Mean F.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Conversations</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0.62</td>
</tr>
<tr>
<td>(Items 1-7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Talks (Items 8-15)</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0</strong></td>
<td><strong>7</strong></td>
<td><strong>4</strong></td>
<td><strong>2</strong></td>
<td><strong>2</strong></td>
<td><strong>0</strong></td>
<td><strong>0.53</strong></td>
</tr>
</tbody>
</table>

The mean F.V. of Part I in the listening test was 0.53, which was lower than the F.V.s of Part II and Part III (0.58 and 0.55, respectively). As shown in Table 15, Part I had more 'slightly too difficult' items (seven items in total, and the F.V.s ranged between 0.25 and 0.45) and only two 'moderately easy' items (whose F.V.s ranged between 0.60 and 0.75). The F.V. results did not meet the expectation that item types should be arranged in order of ascending difficulty. Therefore, the research team further explored the factors affecting examinees' performance and the ways of improving the test content and format in the hope of increasing the validity of this part.

As for test content, a variety of topics were used. Items 1-7 were short conversations in different contexts. Items 8-15 were short talks, such as advertisements, short narrations, speeches, news reports, etc., and the topics included humanities, arts, environment, biology, physical education, etc. The item design aimed at fully orienting the test content with descriptors of Advanced-level listening ability by providing diverse and balanced topics. The amount of time provided for responding to each question was ten seconds. Questions were recorded, and only the choices were printed in the test booklet in order to reduce the probability that examinees would guess items correctly by only reading the questions printed in the test booklet. Also, examinees had to choose the answer in real time. Some of the test

---

6 The top 27% of the test scorers are categorized as the high scoring group (H) and the bottom 27% as the low scoring group (L).

7 According to the survey results, 63% of the examinees considered the ten-second period to be just right for responding to each question, but 36% slightly too short.
takers might not have been able to comprehend short talks with varied contexts and topics in such a short period of time, and therefore, did not perform well on the test, which was particularly evident in responses to items 8-15 (Short Talks). These eight items covered short talks of different types, topics, and contexts, and five of them were 'slightly too difficult.' The variance in these eight items contributed to a lower mean F.V., 0.45. By contrast, items 1 to 7 were all conversations and differed only in contexts. Therefore, items 1-7 achieved a higher mean F.V., 0.62., and just two items were 'slightly too difficult.'

To solve the above low F.V. problem, the research team suggested that a short fronting be added to each conversation/talk, providing information on the types, topics, or contexts of the conversation/talk concerned. Take items 6 and 11 in the listening pretest as examples. The former is a business conversation, and the latter a theatrical commentary broadcast on the radio. The F.V.s were 0.37 and 0.35, respectively. These two items were both slightly too difficult. The following italicized frontings might be provided in advance of the conversation/talk.

6. **Listen to a conversation in which two business people are negotiating an order.**

   M: Now about the CE-10 microphones…I'd like to order 1,000 units but only if you can deliver them by the end of April.
   W: I'm afraid that doesn't give us enough time. We've received several other orders already. What about the first week of May?
   M: Let's make this arrangement: if the shipment arrives *after* the end of April, the price will be discounted 5%.
   W: All right. It's a deal.

   Question: Why might the shipment *not* be delivered by the end of April?

11. **Listen to part of a radio program in which a man is reviewing a theatrical performance.**

   Last night was the opening performance of Shakespeare's *Hamlet* at the University Theater, directed by Robert Ray. The set design, above all, was really impressive, and Ray makes use of the whole theater, with actors and actresses sometimes walking down into the audience. The senior playing Hamlet, Curtis Lamb, did an excellent job; some of the other roles were a bit weak, but perhaps that was just first night nerves.

   Question: What is the speaker's attitude towards this production?

   Buck's study (1991) suggested that although frontings did not have a significant impact on test performance, they helped ease pressure from taking a listening test and had a positive influence on examinees' sentiments.

   As for multiple-choice items, item 15 achieved a D.I. as low as 0.20, which was also the only item whose D.I. was below 0.20. The test content and statistical performance results are listed below:

LR-18
15. What would you say to a tennis racquet that weighed less than your morning cup of coffee? The Prince Thunder titanium racquet—weighing just 7 ounces—is so much lighter than other racquets that it's hard to believe it's standard size. It's also the stiffest racquet you can buy, so it adds force to your stroke while allowing you to control the ball more effectively.

Question: What should we expect from the Prince Thunder tennis racquet?

<table>
<thead>
<tr>
<th>Option</th>
<th>L</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Smaller size.</td>
<td>24%</td>
<td>22%</td>
<td>25%</td>
</tr>
<tr>
<td>B. Greater weight.</td>
<td>25%</td>
<td>3%</td>
<td>14%</td>
</tr>
<tr>
<td>C. Increased shock absorption.</td>
<td>25%</td>
<td>38%</td>
<td>31%</td>
</tr>
<tr>
<td>D. More power.</td>
<td>25%</td>
<td>37%</td>
<td>30%</td>
</tr>
</tbody>
</table>

The main reason behind the low D.I. might be that choice C was longer than the other choices and had a different structure from the others. As a result, more examinees of both the high scoring group and the low scoring group chose C rather than the correct choice D. If choice C was changed to 'Less control', each of the four choices would contain two words in a comparative form. In this way, none of the three choices would cause too much distraction and affect the D.I. value.

2. Part II (Long Conversations)

Table 16  Distribution of F.V.s of Part II in the listening test

<table>
<thead>
<tr>
<th>Sections/Items</th>
<th>Difficulty</th>
<th>Too difficult</th>
<th>Slightly too difficult</th>
<th>Moderately difficult</th>
<th>Moderately easy</th>
<th>Slightly too easy</th>
<th>Too easy</th>
<th>Mean F.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation 1 (Items 16-20)</td>
<td>&lt;0.25</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0.64</td>
</tr>
<tr>
<td>Conversation 2 (Items 21-27)</td>
<td>0.25-0.45</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.53</td>
</tr>
<tr>
<td>Total</td>
<td>0.46-0.59</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60-0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.76-0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean F.V. of Part II was 0.58, which was the highest of the three parts in the listening pretest. As shown in Table 16, most of the Part II items were moderately difficult or moderately easy, with F.V.s ranging between 0.46 and 0.90. The difficulty distribution of Part II items was better than that of Part I items.

As for the test content, Part II consisted of two long conversations differing in topics and item type design. The first conversation, an interview with a best-seller author, was less abstract, and the item type was short answer questions aimed at assessing whether examinees could grasp detailed information on the author's writing career (for example, 'When did Taylor
realize he loved to write? What does Taylor prefer to call the kind of book he writes?

The second conversation, where two persons discussed an article about migration and talked about
the nature and formation of different migration types, was more abstract, and the item type,
notes-completion, aimed at assessing whether examinees could grasp the topic and structure
of the talk. The second conversation tested global comprehension. As for the test format, the
questions in this part were printed in the test booklet so examinees could know in advance
which part of the conversations they should pay attention to. As shown in Table 16, the mean
F.V. of the first conversation items (items 16-20) was 0.64, higher than that of the second
conversation items (items 21-27), 0.53. The research team speculated that the two following
variables would correlate with test performance in long conversations.

1) **The degree of abstraction/concreteness of topics**: Everyday life topics were more concrete
and less difficult compared with academic or abstract topics.

2) **The focus of questions**: Examinees seemed to better grasp detailed information in long
conversations when they knew the focus of questions in advance of listening to the
conversations. While comprehension of conversational structures involved greater ability to
integrate the information heard, examinees generally did not perform well in this regard.

### 3. Part III (Long Talks)

**Table 17 Distribution of F.V.s of Part III in the listening test**

<table>
<thead>
<tr>
<th>Sections/Items</th>
<th>Too difficult</th>
<th>Slightly too difficult</th>
<th>Moderately difficult</th>
<th>Moderately easy</th>
<th>Slightly too easy</th>
<th>Too easy</th>
<th>Mean F.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long Talk 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Items 28-34)</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td>Long Talk 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Items 35-40)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1</strong></td>
<td><strong>2</strong></td>
<td><strong>3</strong></td>
<td><strong>5</strong></td>
<td><strong>2</strong></td>
<td><strong>0</strong></td>
<td><strong>0.55</strong></td>
</tr>
</tbody>
</table>

The mean F.V. of Part III was 0.55, falling between those of the other two parts in the
listening test. As shown in Table 17, most of the Part III items were moderately difficult or
moderately easy, with the F.V.s ranging between 0.46 and 0.75. The item difficulty of this part
was generally satisfactory.

As for the topics, the first talk was a speech on how to reform the healthcare
environment, while the second one was a news report on the rampant piracy at sea in recent
years. Both of the talks were equally abstract. Although the two talks used different item types
(the first was short answer questions, and the latter notes-completion), both of them contained
two items testing 'global comprehension' and four or five items testing 'local comprehension'.

LR-20
The distribution of the test focuses was similar. As shown in Table 17, the mean F.V.s of both long talks were 0.55, meaning the two talks were of equal difficulty. It was speculated by the research team that the same F.V. results might be attributed to the consistency of the abstract level of topics and normal distribution of item focuses between the two talks.

As for non-multiple-choice items, item 39 of this part achieved the lowest F.V., 0.17 and an acceptable D.I., 0.34. With the lowest F.V., item 39, one of the second long talk items, was the most difficult. Its test content and statistical results are shown as below:

**Two reasons why shipping companies don't report attacks to the police:**

1. Higher (39)…………………………………. would result.
2. Stolen ships and cargo are difficult to recover.

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Answer</td>
<td>5%</td>
<td>37%</td>
<td>17%</td>
</tr>
<tr>
<td>Wrong Answer</td>
<td>80%</td>
<td>63%</td>
<td>78%</td>
</tr>
<tr>
<td>Unanswered</td>
<td>15%</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>F.V.=0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.I.=0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item 39 tested one of the reasons why shipping companies did not report attacks to the police, and the answer was 'the cost of insurance (or insurance cost) and wages.' The key words 'the cost of insurance' and 'wages' appeared in the latter part of the news report, 'The companies are worried that if they report such incidences, the cost of insurance will increase and crewmembers will demand higher wages.' As the item did not specify the key answer included two parts, many of the examinees wrote either 'the cost of insurance' or 'wages'. Generally, such incomplete answers were not given any scores. It was believed that the low F.V. of item 39 was caused by the unclear instruction of the item. Therefore, the research team suggested that such multi-part items (meaning a key answer consisted of two or several parts) should be avoided in designing short answer and notes-completion questions. Take item 39 as an example. The notes-completion question might be revised to "Higher (39)…………………… and wages would result." In this way, the answer would be "cost of insurance (or insurance cost)" only, which could reduce examinees' confusion and save raters' trouble.

As discussed above, by providing conversations and talks on different topics, this listening test was aimed at assessing examinees' listening ability to achieve global comprehension and local comprehension, and make inferences about the heard information. Test performance was found to highly correlate with two factors: 1) the choice of source texts, including contextual features and the concrete/abstract level of topics; 2) the main points of listening questions and the level of comprehension required. In addition, wording of listening questions and lexis and structure of choices were also among the important factors that affected item difficulty of this listening pretest.
IV. Analysis of Sample Reading Pretest Items

Analysis of sample reading pretest items and examinees' performance are presented below. Also, suggestions about how to improve item writing are provided as references for future test compilation. The quantitative and qualitative data analyses of this reading pretest were generally conducted in the same way as the listening pretest.

1. Part I (Careful Reading)

Table 18  Distribution of F.V.s of Part I in the reading test

<table>
<thead>
<tr>
<th>Sections/Items</th>
<th>Difficulty</th>
<th>Too difficult</th>
<th>Slightly too difficult</th>
<th>Moderately difficult</th>
<th>Moderately easy</th>
<th>Slightly too easy</th>
<th>Too easy</th>
<th>Mean F.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1 (Items 1-5)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0.78</td>
</tr>
<tr>
<td>Text 2 (Items 6-9)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.71</td>
</tr>
<tr>
<td>Text 3 (Items 10-13)</td>
<td></td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.36</td>
</tr>
<tr>
<td>Text 4 (Items 14-20)</td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>0.60</td>
</tr>
</tbody>
</table>

The mean F.V. of Part I in the reading pretest was 0.60, higher than that of Part II. As shown in Table 18, half of the items were moderately difficult or moderately easy (the F.V.s ranging from 0.46 to 0.75), another five of the items were slightly too difficult or too difficult (the F.V.s were below 0.45), and the remaining five items were slightly too easy, indicating there was a wide distribution of item difficulty.

Part I in the reading test consisted of four texts, and the F.V. results showed that test items were arranged in order of ascending difficulty. The first and second texts were simpler, with the mean F.V.s achieving 0.78 and 0.71, respectively. The topics were a new medical discovery and eco-system protection, respectively. The fourth article, which concerned perfume, had a mean F.V. of 0.53. The mean F.V.s of the items under Text 4 ranged within acceptable bounds of 0.24 and 0.77.

It was noteworthy that Text 3 (items 10-13), which is about Nigerian writer Chinua Achebe's motivation for becoming a writer, his writing style, and well-known works, achieved a low mean F.V., 0.36. More specifically, the F.V. of item 13 was as low as 0.15, and items 10 and 12 did not discriminate well, with the D.I.s being 0.16 and 0.19, respectively. Based on the analysis, the low F.V. and D.I. problems were probably due to unclear wording of some of the questions and insufficient key information provided in the text. Take item 12 as an example. Its content and statistical results are listed below:
12. How does Achebe present Europeans in his novels?

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. As superior to blacks in most ways</td>
<td>34%</td>
<td>27%</td>
<td>24%</td>
</tr>
<tr>
<td>B. As inferior to blacks in most ways</td>
<td>15%</td>
<td>10%</td>
<td>22%</td>
</tr>
<tr>
<td>C. As gentle but ambitious</td>
<td>28%</td>
<td>14%</td>
<td>20%</td>
</tr>
<tr>
<td>D. As benevolent but interfering</td>
<td>22%</td>
<td>49%</td>
<td>34%</td>
</tr>
</tbody>
</table>

Key=D

Both the F.V. and D.I. of Item 12 were very low (0.34 and 0.19, respectively). There might be two reasons behind this. First, the word "Europeans" in the question was not clear or specific enough. It might add clarity to the question if it were revised as "How does Achebe present the British colonists in his novels?" Second, the key to this question was D: As benevolent but interfering; however, the information provided in the text was not sufficient for examinees to infer that the British colonists were "benevolent," which might explain why some strong examinees did not choose D. The last part of paragraph five, "The old order is shattered. It cannot withstand the challenge of the new," and the middle part of paragraph eight, "Once again, the conflict is between Ibo and the English, tradition and modernity," showed that the British colonists introduced modernity from the West to African tribes but also caused the breakdown of local cultural traditions. Therefore, choice D might be rephrased as "As modernizing but interfering."

To solve the problem of unclear key information, item writers could do "text-mapping" in the process of item writing, as suggested by Urquhart & Weir (1998) and Weir et al. (2000). Basically, "text-mapping" involves two important steps: 1) At the beginning of item writing, item writers have to collect sufficient key information on a topic and make it the basic structure of a text; 2) a trial test has to be conducted after the item writing is finished. Potential examinees might be testing professionals who were not involved in the item writing or people whose English proficiency reached the GEPT Advanced level. These examinees are asked to specify the main ideas and important details of the text. The results are compared with item writers’ opinions to seek an objective consensus.

As for non-multiple-choice items, the research team found that lexical overlap between a question and its related text had an influence on test performance. Take item 7 under Text 2 on eco-system protection as an example. Its content and statistical results are listed below:

7. How is the sixth extinction unlike earlier ones?

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Answer</td>
<td>78%</td>
<td>94%</td>
<td>88%</td>
</tr>
<tr>
<td>Wong Answer</td>
<td>22%</td>
<td>6%</td>
<td>12%</td>
</tr>
<tr>
<td>Unanswered</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

F.V.=0.88  D.I.=0.20
The F.V. and D.I. of this item were 0.88 and 0.20, respectively, indicating that the item was slightly too easy and did not discriminate well. The word "sixth" appeared in the sentence "The present mass extinction, the sixth, is different, though, from all of its predecessors—not just in "the raging speed" at which it is proceeding but because, for the first time, it was unleashed by mankind." As observed by the research team, examinees tended to make good use of numerical clues. They might not fully understand the meaning of the sixth extinction, but they could still get the item right with the number "six" as a clue. Therefore, up to 78% of the low scoring examinees got the item right. To improve the low D.I., the question might be rephrased as "How is the current extinction unlike earlier ones?"

2. Part II (Skimming & Scanning)

Table 19  Distribution of F.V.s of Part II in the reading test

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Too difficult</th>
<th>Slightly too difficult</th>
<th>Moderately difficult</th>
<th>Moderately easy</th>
<th>Slightly too easy</th>
<th>Too easy</th>
<th>Mean F.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sections/Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 2A(Items 21-26)</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0.67</td>
</tr>
<tr>
<td>Part 2B (Items 27-32)</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.46</td>
</tr>
<tr>
<td>Part 2C(Items 33-40)</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0</strong></td>
<td><strong>4</strong></td>
<td><strong>11</strong></td>
<td><strong>3</strong></td>
<td><strong>2</strong></td>
<td><strong>0</strong></td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>

The mean F.V. of Part II was 0.54, lower than that of Part I (0.60). As shown in Table 19, most of the Part II items were moderately difficult (11 items in total), achieving F.V.s between 0.46 and 0.59, while only three of the items were moderately easy. The F.V. distribution shown in Table 19 could be deemed reasonable as most of the examinees were much more familiar with the item type in Part I (Careful Reading) than that in Part II (Skimming & Scanning).

However, the D.I. analysis showed that two of the six Part 2A (main ideas of passages) items were below 0.20, while the items in Part 2B (main ideas of paragraphs) and Part 2C (looking up specific information) generally discriminated well. Furthermore, it was found that the most examinees read texts word by word and spent more than ten minutes answering items 21-26, which was half of the time allowed for finishing Part II. Therefore, it was necessary to review the test content and test format of Part 2A to explore the factors that might influence test performance.

Part 2A consisted of four passages. Each of the passages contained 150-300 words and was preceded by one or two related multiple-choice questions that tested global comprehension by asking about the topics, main ideas, or purposes of the passages. Examinees had to read questions first and find key information through a related passage. As indicated by the test
performance and our consultants' opinions, the short length of the Part 2A passages and the multiple-choice item type, which was familiar to the examinees, invited the examinees to read the passages word by word.

Based on the above analysis, it was suggested that the length of Part 2A passages be extended in the future. Part 2B (main ideas of paragraphs/heading matching), which was meant to be completed with a skimming strategy, achieved good F.V. and D.I. results. Therefore, the research team suggested that the short passages in Part 2A be replaced by a long text of 600-800 words, and the item type be changed from multiple-choice questions to matching. Also, as narrative and descriptive texts were used in Part 2B and Part 2C, it was suggested that an argumentative text be added in Part II for a balance of text types. Table 20 shows a revised design of Part II item types.

Table 20  Design of Part II item type in the reading test (revised version)

<table>
<thead>
<tr>
<th>Part</th>
<th>Section</th>
<th>Test focus</th>
<th>Text type/length</th>
<th>Item type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part II:</td>
<td>Part 2A</td>
<td>Skimming</td>
<td>One narrative text and one argumentative text</td>
<td>Heading matching</td>
</tr>
<tr>
<td>Skimming &amp;</td>
<td></td>
<td></td>
<td>(each contained 600-800 words)</td>
<td></td>
</tr>
<tr>
<td>Scanning</td>
<td>Part 2B</td>
<td>Scanning</td>
<td>Three interrelated descriptive or expository</td>
<td>Notes completion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>texts (each contained approximately 300 words)</td>
<td></td>
</tr>
</tbody>
</table>

3. Factor analysis of the reading test results

With reference to previous studies, the reading test mainly featured the separation of Careful Reading and Skimming & Scanning due to the difference of the tested constructs. The two individual sections also differed in the test focus, required reading rate, item format, and the amount of time provided for responding to questions. As shown in the above analyses of item difficulty and correlation coefficients, the mean F.V. of Skimming & Scanning, 0.54, was lower than that of Careful Reading, 0.60, and the correlation between the two was 0.61, indicating the abilities tested were relevant to each other, but not identical.

The above results showed a certain degree of heterogeneity in constructs between the two individual sections. To further analyze the heterogeneity, factor analysis was conducted on the test performance of the 234 examinees. More specifically, the results of principal component analysis showed that the eigenvalue of factor 1 was larger than one. The eigenvalue of factor 2 was smaller than one (0.92), but the factor could still be retained according to the scree test result. The factor matrix analysis results of the two factors are shown in Table 21.
Table 21  Summary of factor analysis results (unrotated)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Careful Reading (Part I)</td>
<td>0.84</td>
<td>0.08</td>
</tr>
<tr>
<td>Skimming &amp; Scanning (Part II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main ideas of essays (previously Part 2A)</td>
<td>0.64</td>
<td>-0.58</td>
</tr>
<tr>
<td>Main ideas of paragraphs (Previously Part 2B)</td>
<td>0.76</td>
<td>-0.16</td>
</tr>
<tr>
<td>Looking up specific information (Previously Part 2C)</td>
<td>0.57</td>
<td>0.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Variance %</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.02</td>
<td>50.59%</td>
<td>50.59%</td>
</tr>
<tr>
<td>0.92</td>
<td>22.95%</td>
<td>73.54%</td>
</tr>
</tbody>
</table>

To meet the "simple structure" principle, the factor loading of some test sections under factor 1 was made as large as possible while that under the other common factors was made close to zero in order to facilitate the identification and naming of the common factors. Table 22 shows the results of varimax-rotation.

Table 22  Summary of factor analysis results (orthogonal rotation)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Careful Reading (Part I)</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>Skimming &amp; Scanning (Part II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main ideas of essays (previously Part 2A)</td>
<td>0.85</td>
<td>-0.10</td>
</tr>
<tr>
<td>Main ideas of paragraphs (Previously Part 2B)</td>
<td>0.71</td>
<td>0.30</td>
</tr>
<tr>
<td>Looking up specific information (Previously Part 2C)</td>
<td>0.04</td>
<td>0.93</td>
</tr>
</tbody>
</table>

As shown in Table 22, 1) Part 2A (Main ideas of essays) and Part 2B (Main ideas of paragraphs) had greater factor loading under factor 1; 2) Part 2C (Looking up specific information) had greater factor loading under factor 2; 3) Careful reading had great factor loading under both factors 1 and 2. Factor 1 was considered to be global comprehension ability, while factor 2 might be scanning ability. However, as mentioned above, many of the examinees spent too much time reading the Part II texts word by word and did not use scanning strategies until they found they were running out of time. The overuse of careful reading strategies might have affected the construct validity of Skimming & Scanning, thus causing Careful Reading, Part 2A, and Part 2B to have greater factor loading under factor 1.
V. Conclusion and Suggestions

The overall F.V.s of the listening and reading pretests were 0.55 and 0.57, respectively. The F.V. values were a bit lower than the expectations of the research team but were still deemed reasonable. As for the D.I. results, more than 85% of the items in the listening and reading tests discriminated well. The overall reliability of the two tests (eighty items in total) was 0.91; the overall reliability of the listening test (forty items in total) was 0.88, while that of the reading test (forty items in total) was 0.83. All of the reliability indices were higher than 0.80. The overall reliability of the reading test was a bit lower than that of the listening test, which was considered to correlate with the heterogeneity of the construct validity between Careful Reading and Skimming & Scanning.

As planned for the GEPT Advanced Level tests, the cut score of the listening and reading tests were 80 (i.e. correctly answering twenty-seven or more items) out of a total score of 120, respectively. By that standard, fifty-eight examinees passed the first stage of the advanced level tests, accounting for 25% of all examinees. If examined by pretest groups, 30% of university seniors majoring in English passed the first stage of the advanced level tests. 33% of those who had passed the second stage of the GEPT high-intermediate test passed the first stage of the advanced level tests. Forty-eight examinees who only passed the first stage of the HI test were also sampled, but none of them passed the first stage of the advanced level tests.

The results of the above statistical analysis results generally met the expectations of the research team, indicating that this pretest could serve as reference for future advanced level test compilation. However, the following modifications were suggested for future tests:

1) In the listening test, the mean F.V. of Part I (Short Conversation & Talk), 0.53, was lower than that of Part II and Part III, which ran counter to the original expectation that Part I should have functioned as a warm-up. It was suggested that a short fronting be added to each conversation/talk, providing information on the types, topics, or contexts of the conversation/talk concerned in order to solve the low F.V. problem of Part I.

2) In the reading test, the mean F.V. of Part II (Skimming & Scanning), 0.54, was lower than that of Part I (Careful Reading), 0.60, which might result from the examinees' unfamiliarity with the item type of skimming and scanning and overuse of careful reading strategies. Further analysis of Part II item difficulty showed that Part 2A achieved a mean F.V. of 0.67, higher than the other sections in Part II. It is thought that the short passages (each contained 150-300 words, with a total length of approximately 900 words) and the multiple-choice questions in Part 2A invited the examinees to read carefully. Therefore, in the future, short essays would be altogether replaced by an extended argumentative text of 600-800 words, and the original multiple-choice questions would be replaced by matching.
3) This advanced level listening and reading tests were the first GEPT test that used non-multiple-choice item types of notes completion and short answer questions to assess reading and listening. The pretest results showed that most of the non-multiple-choice items had reasonable F.V.s and D.I.s. However, item writers had to pay special attention to the following factors that might influence test performance: ① the wording of questions should be clear, and multi-part items should be avoided; ② the key information should be sufficient and clear; ③ the degree of lexical overlap between a question and its related text should be appropriate.

In addition, the item type, Skimming & Scanning, was first used in this advanced level reading test. Although the examinees had been provided with detailed instructions for the test method and notified of the test duration, many of them still read the texts word by word. Therefore, more detailed instructions and sample items should be provided in advance of the operational tests to help examinees become familiar with this test format, which might lead to a positive washback effect on the teaching of English reading in Taiwan.
Appendix 1: The GEPT Advanced Level Listening Pretest form

GENERAL ENGLISH PROFICIENCY TEST

ADVANCED LISTENING COMPREHENSION TEST

This listening comprehension test will test your ability to understand spoken English. There are three parts to this test. Special instructions will be given to you at the beginning of each part.

Part I

In this part of the test, you will hear 15 short conversations and talks. At the end of each conversation and talk, you will be asked a question about what was said. You will hear each conversation, talk and question just one time. You may take notes in your test booklet as you listen. After you hear the question, read the four choices in your test booklet and choose the best answer to the question you have heard.
1. A. They're poor housekeepers.
   B. They just moved into a house.
   C. They're cleaning the house before guests arrive.
   D. They just returned home from a vacation.

2. A. Express her opinions more clearly.
   B. Make another copy of her paper.
   C. Indicate her sources in the paper.
   D. Correct some grammatical errors.

3. A. They're planning a trip to Switzerland.
   B. They wish they could travel like Ruth.
   C. They don't think Ruth is responsible.
   D. They feel Ruth shouldn't have left.

4. A. Stay away from Tim.
   B. Forgive Tim's rudeness.
   C. Ask for Tim's permission.
   D. Challenge Tim about his behavior.

5. A. At the end of the second term.
   B. After February 21.
   C. On May 21.
   D. In the next two days.

6. A. The company is busy with earlier orders.
   B. The customer ordered more than 1,000 units.
   C. The microphone is out of stock now.
   D. The customer is getting a discount.

7. A. She'd rather have watched another program.
   B. She didn't finish watching the movie.
   C. It was an exciting movie.
   D. It was based on a real life story.

8. A. Why dancing was added to caroling.
   B. How the meaning of the word "carol" changed.
   C. The spread of Christmas carols internationally.
   D. The function of carols today.

9. A. It lowered global temperatures.
   B. It led to another volcanic eruption.
   C. It enlarged the holes in the ozone layer.
   D. It caused droughts in both hemispheres.

10. A. The rain poured in the windows.
    B. Water rushed in under the door.
    C. Water came through the ceiling.
    D. The pipes burst, flooding the apartment.

11. A. He feels the actors weren't directed well.
    B. He's impressed by the group's fine acting.
    C. He likes everything except the sets.
    D. He's satisfied with the overall effect.

    B. Adult literacy campaigns.
    C. The plight of aboriginal people.
    D. The battle against terrorism.

13. A. The unpleasantness for the mother.
    B. The mother's nesting habits.
    C. The large number of eggs.
    D. The father's role at the birth.

14. A. Purchasing more high-quality objects.
    B. Organizing its museum collections.
    C. Protecting natural preserves.
    D. Building a new cultural center.

15. A. Smaller size.
    B. Greater weight.
    C. Increased shock absorption.
    D. More power.
Part II

In this part of the test, you will hear two long conversations. You will hear each of them just one time. You may take notes in your test booklet as you listen.

Before you listen to each conversation, you will have 30 seconds to read the instructions and questions which are printed in your test booklet.

After you hear each conversation, you will have 3 to 4 minutes to answer the questions. Please write your answers on your answer sheet. Each answer will require just a word, phrase or short sentence.
Questions 16-20

Listen to the following interview with a writer of horror novels and answer each question with a word, phrase or short sentence.

16. When did Taylor realize that he loved to write?

17. Did Taylor's first book come out before, during, or after he worked in advertising?

18. What does Taylor prefer to call the kind of book he writes?

19. In Taylor's opinion, why do people read horror novels?

20. Which would be the best title for Taylor's next book?
   A. Fright Masters  
   B. Final Trap  
   C. Wasted Years  
   D. Black Death
Part II continues.

Do NOT turn this page until you are instructed to do so.
Questions 21-27

Listen to the following discussion between a man and a woman about a subject which the woman has been reading about and then fill in the missing information. Each answer will require just a word, phrase or short sentence.

The **overall subject** of the article Carol has been reading: (21)……………………………

The **four specific types of people** Carol mentioned:

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I:</td>
<td>35 million</td>
</tr>
<tr>
<td>Type II:</td>
<td>30 million</td>
</tr>
<tr>
<td>Type III:</td>
<td>60 million</td>
</tr>
<tr>
<td>Type IV:</td>
<td>80 million</td>
</tr>
</tbody>
</table>

- **Type I:** (22)……………………………………………………
- **Type II:** (23)………………………………………………
- **Type III:** People who (24)………………………………
  - because (25)………………………………………………
- **Type IV:** People from (26)………………………………
  - relocating in (27)……………………………………

LR-34
Part III

In this part of the test, you will hear two long talks. You will hear each of them just one time. You may take notes in your test booklet as you listen.

Before you listen to each talk, you will have 30 seconds to read the instructions and questions which are printed in your test booklet.

After you hear each talk, you will have 3 to 4 minutes to answer the questions. Please write your answers on your answer sheet. Each answer will require just a word, phrase or short sentence.
Questions 28-34

Listen to part of a lecture given at a conference and answer each question with a word, phrase or short sentence.

Hospital Innovation

✧ According to the speaker, what did hospitals neglect to consider in the past?

(28)............................................................................................................

✧ What is the first major innovation that the speaker mentions?

(29)............................................................................................................

✦ How could new hospital buildings implement this?

(30)............................................................................................................

✦ What three alternative ways does the speaker mention for implementing this?

(31)............................................................................................................

(32)............................................................................................................

(33)............................................................................................................

✧ What is the second major innovation that the speaker mentions?

(34)............................................................................................................
Part III continues.

Do NOT turn this page until you are instructed to do so.
Questions 35-40

Listen to the following news report about modern-day pirates and then fill in the missing information. Each answer will require just a word, phrase or short sentence.

Modern-day Pirates

✧ Two trends in modern-day piracy:
  1. The number of attacks is (35).................................................................
  2. Pirates are becoming (36)........................................................................

✧ High risk areas:
  1. Harbors in Brazil because (37)..............................................................
  2. South East Asian waters because (38).....................................................

✧ Two reasons why shipping companies don't report attacks to the police:
  1. Higher (39)............................................................................................. would result.
  2. Stolen ships and cargo are difficult to recover.

✧ Two main purposes of the Piracy Center:
  1. To help ships report both attacks and attempted attacks
  2. (40)........................................................................................................

-The End-
The GEPT Advanced Level Listening Pretest Script:

GENERAL ENGLISH PROFICIENCY TEST
ADVANCED LISTENING COMPREHENSION TEST

This listening comprehension test will test your ability to understand spoken English. There are three parts to this test. Special instructions will be given to you at the beginning of each part.

Part I
In this part of the test, you will hear 15 short conversations and talks. At the end of each conversation and talk, you will be asked a question about what was said. You will hear each conversation, talk and question just one time. You may take notes in your test booklet as you listen. After you hear the question, read the four choices in your test booklet and choose the best answer to the question you have heard.

Please turn to the next page. [pause 4 seconds] Now we will begin Part I with question no. 1.

1. W: Bruce, come look at this!
   M: What's the matter?
   W: I tried to open these drapes, and this one just fell off the rod! They're literally rotten!
   M: That doesn't surprise me. I just discovered that at least half of the outdoor lights don't work!
   W: The people who lived here before certainly didn't take very good care of this place, huh?

   Question: What can we infer about the speakers?

2. W: Thanks for taking the time to look at the first draft of my paper.
   M: That's okay. You said you didn't have much experience with research papers?
   W: Very little. As an undergraduate, I mostly did group projects. So what do you think?
   M: (sighs) Sandy, I'm afraid you've got some serious problems here.
   W: Really?
   M: For example, I've marked several pages where you apparently copied some other source, but you didn't cite it.

   Question: According to the man, what will the woman need to do?

3. M: Hey, have you heard from Ruth lately?
   W: I just got an e-mail yesterday. She's in Switzerland!
   M: Switzerland! You mean she's still in Europe? How long has she been gone?
   W: Eight months, so far.
   M: Eight months! Gosh, wouldn't it be nice to have no job responsibilities, and be able to see the world?
   W: Would it ever!

   Question: What can we infer about the speakers?
4. M: Are you going to call Tim or should I?
   W: I think you should. I always get cold feet about confrontations.
   M: Well, we're in the right. He shouldn't take things without asking permission first.
   W: I know but…
   M: Listen, if we don't get some backbone now, he'll just keep walking all over us.

**Question: What does the man want to do?**

5. M: Hello?
   W: Yes, I'm calling about the business writing course.
   M: What would you like to know?
   W: Is there still space available for the second term?
   M: Yes, there is. But you'd better hurry. Registration ends on February 21.
   W: February 21…that's the day after tomorrow.
   M: Yes, that's right. Or if you don't mind waiting until the third term, you can register before the end of May.
   W: Well, I'm going to New Zealand this summer, so I'll try to register for the second term.

**Question: When will the woman register?**

6. M: Now about the CE-10 microphones…I'd like to order 1,000 units but only if you can deliver them by the end of April.
   W: I'm afraid that doesn't give us enough time. We've received several other orders already. What about the first week of May?
   M: Let's make this arrangement: if the shipment arrives after the end of April, the price will be discounted 5%.
   W: All right. It's a deal.

**Question: Why might the shipment not be delivered by the end of April?**

7. M: Hi, Linda. Did you watch the baseball game on Channel 15 last night?
   W: No, Kevin wanted to watch an adventure movie on Channel 40, so I watched it with him.
   M: So, how was it?
   W: Believe it or not, I was on the edge of my seat during the whole movie.

**Question: What does the woman imply about last night's movie?**

8. (W) As we continue studying the origins of Christmas customs, let's look at Christmas carols. All over the world today, people sing special songs about the birth of Jesus, and groups may even stand on street corners or walk from house to house singing these so-called Christmas carols. What you may not realize is that the word "carol" originally referred to an English folk dance which included singing.

**Question: What will the speaker probably talk about next?**
9. (M) Before Mt. Pinatubo in the Phillipines erupted in 1991, scientists had issued warnings that the earth was getting warmer, due to the depletion of ozone in the atmosphere. The volcano's eruption, however, led to some global cooling in the following few years. The ash and sulfur-dioxide gas spewed into the atmosphere by the eruption created a haze which reflected sunlight, thus reducing the temperature on the earth's surface. This dust cloud diffused and covered both hemispheres.

Question: According to this talk, what effect did the eruption of Mt. Pinatubo have on the earth?

10. (W) I once lived in India, in a tiny top-story apartment. The roof of our building was made of tin sheets, and the ceiling below was wooden boards. One winter, there was a terrible storm, with extremely strong winds. The wind and rain beat hard against the metal roof for hours. Then one blast of wind began to push up the eastern side of the roof. Soon we heard a ripping sound, and some of the metal sheets slid off the roof, crashing into the road below.

Question: What probably happened next?

11. (M) Last night was the opening performance of Shakespeare's *Hamlet* at the University Theater, directed by Robert Ray. The set design, above all, was really impressive, and Ray makes use of the whole theater, with actors and actresses sometimes walking down into the audience. The senior playing Hamlet, Curtis Lamb, did an excellent job; some of the other roles were a bit weak, but perhaps that was just first night nerves.

Question: What is the speaker's attitude towards this production?

12. (W) A native of Guatemala and an indigenous person herself, Rigoberta Menchu has emerged as a spokesperson for indigenous peoples around the world. Through her writings, she has called attention to the danger these groups are in—the danger of disappearing altogether, along with their native languages and cultures. Menchu was awarded the Nobel Peace Prize for her efforts in 1992.

Question: According to this talk, what subject has Rigoberta Menchu mainly written about?

13. (M) Another interesting thing about the kiwi bird is that it has one of the largest eggs of any bird. In fact, it's much larger than you would expect for a bird of that size. The egg becomes so large, and fills the female's body so completely, that there's no room for food. The bird sometimes soaks itself in cold rain water to relieve the discomfort.

Question: What does the speaker emphasize about the birth of kiwi birds?

14. (W) The U.S. National Park Service administers more than 200 cultural and historic sites, including Independence Hall, the Statue of Liberty, and 150 nature preserves. In addition, the Park Service storerooms hold more than 25 million museum-quality objects. However, vast parts of the collection remain uncatalogued, due to a shortage of funds, and many of the artifacts are rotting away.

Question: What does the National Park Service need more money for?

15. (M) What would you say to a tennis racquet that weighed less than your morning cup of coffee? The Prince Thunder titanium racquet—weighing just 7 ounces—is so much lighter than other racquets that it's hard to believe it's standard size. It's also the stiffest racquet you can buy, so it adds force to your stroke while allowing you to control the ball more effectively.

Question: What should we expect from the Prince Thunder tennis racquet?
Part II

In this part of the test, you will hear two long conversations. You will hear each of them just one time. You may take notes in your test booklet as you listen.

Before you listen to each conversation, you will have 30 seconds to read the instructions and questions which are printed in your test booklet.

After you hear each conversation, you will have 3–4 minutes to answer the questions. Please write your answers on your answer sheet. Each answer will require just a word, phrase or short sentence. Please pay attention to grammar and spelling. Answers which cannot be understood will be considered incorrect.

Please turn to the next page. [Pause 4 seconds] Now we will begin Part II with questions no. 16 to 20.

For questions no. 16 to 20, you will listen to an interview with a writer of horror novels. Before listening, you will have 30 seconds to read the instructions and questions no. 16 to 20. Please begin now. When you finish, look up from your desk.

[Pause 30 seconds.]

Now listen to the interview. You may take notes in your test booklet as you listen.

W: Good morning, and welcome to "Reader's Gallery." Today we have a guest who has written 26 novels, and has sold an amazing 60 million copies worldwide. His latest novel "Black Death" has recently become the best-selling novel in America. We welcome Mike Taylor to our show. Thanks for coming, Mike.

M: It's a pleasure, Laura.

W: Well, we all know Mike Taylor--the writer of best-selling horror novels, but tell us something about Mike Taylor--the person.

M: Okay. Well, I grew up in Los Angeles, California, where I graduated from high school. After that, I went to university to study psychology. I chose to major in psychology because I've always been interested in the mind and how people think, even as a child. I really enjoyed my time there. In fact, it was at university that I discovered my true passion--writing.

W: You wrote for the university's magazine, didn't you?

M: Yes, that's correct. It started out really as a bit of fun, but when my first short story appeared in the magazine, I was over the moon! To be honest, I still feel that way every time my work gets published.

W: So, what did you do after your degree?

M: I began working in an advertising company, but I soon got bored with that. I think I only stayed 7 months! It was afterwards that I decided to be a full-time writer, and so I devoted my time to writing. Then in 1988, my first novel was published.

W: Oh, yes. I remember the book "House of Horror" very well. It became a national best-seller, and it was even made into a very successful Hollywood film. But tell me Mike, why did you choose to write horror stories?

M: Well, I don't actually see myself as a writer of "horror" books. I see myself as a writer of psychological thrillers. I like to touch on real fears, such as the fear of death, or the unknown, and then I like to use that as a basis for my novels.

W: I have always wondered why people are so fascinated by horror films or novels. Why do you think we like to read about "horror"?

M: Well, it's a mixture of, ah…fear, as I said, and um… curiosity, curiosity about ourselves. Everyone is frightened by something, and I think that by exposing ourselves to fear, we are actually trying to understand our own fears.
W: Lots of writers like to try another genre after they've done one sort of novel for awhile. Tom Clancy, for example, is most famous for his stories of international terrorism—spies and so on—but he's also written some non-fiction. Have you thought about…?

M: Doing non-fiction? No, but my next novel will be something of a departure. It will be about fear, but it's not really a thriller—it's about a man who suddenly reaches the end of his life, or almost the end, and he realizes he has hurt all the most important people in his life, everyone he loves. And his fear is that he won't be able to undo this, he won't have any opportunities to heal those relationships before he dies.

W: Do you have a title?

M: Not yet.

W: Well, thank you again for joining us today on "Reader's Gallery," Mike.

M: You're welcome.

Now you will have 3 minutes to answer questions no. 16 to 20. Write your answers on your answer sheet.

[Pause 3 minutes.]

For questions no. 21 to 27, you will listen to a discussion between a man and a woman about a subject which the woman has been reading about.

Before listening, you will have 30 seconds to read the instructions and questions no. 21 to 27. Please begin now. When you finish, look up from your desk.

[Pause 30 seconds.]

Now listen to the discussion. You may take notes in your test booklet as you listen.

(sound effect: flipping the pages)

W: Wow, you'd never believe how many people are moving these days!

M: What do you mean, Carol? Who's moving that we know?

W: I don't mean moving to new apartments. I mean whole populations moving because they have no choice, or very little choice.

M: Oh, I see. You're talking about immigration, aren't you?

W: That's right. I've been reading this article about migration, and believe it or not, there are over thirty-five million of those migrant workers around the world today.

M: Thirty-five million! Wow, that's a lot of people moving!

W: Yes, but those are only the migrants who go to other countries for job opportunities. The article also talked about other types of people migrations. For example, about thirty million people in the world today have been forced to move somewhere else within their own countries. The article called this "internal displacement."

M: Why did they have to move?

W: Sometimes there were wars, or their land was taken by the government. It says here that more than half of those thirty million people are in Africa, and about a quarter are in Asia.

M: Gosh, that's amazing. Imagine so many people having to move.

W: Then there's another category of people who are considered migrants. They're the ones who once made their living from the land—by growing crops or raising animals—but now the environment has been destroyed.
M: Do you mean that the soil has become ruined by poor farming techniques?
W: Yes, or damaged by chemicals. Also, remember that the rainfall patterns have changed, and many areas are becoming as dry as deserts.
M: So there are a lot of migrants who are forced to abandon their land because of environmental destruction.
W: Yeah, around sixty million.
M: So what do they do?
W: Well, some are able to find farm work on other people's land.
M: But I guess many of them try to find work in the cities.
W: Yeah, and they join the *fourth* wave of migration I was reading about—the migrants moving from rural to urban areas. That's been going on for years, even before good agricultural land started to disappear. Apparently there are about eighty million of these migrants right now.
M: Eighty million!
W: And every year another one and a half million are expected to leave the country for the city.
M: Gosh, who's going to be left to grow our food? But I see what you mean now, Carol—there sure are lots of people moving out there!

Now you will have 3 minutes to answer questions no. 21 to 27. Write your answers on your answer sheet.

[Pause 3 minutes.]

Part III
In this part of the test, you will hear two long talks. You will hear each of them just one time. You may take notes in your test booklet as you listen. Before you listen to each talk, you will have 30 seconds to read the instructions and questions which are printed in your test booklet. After you hear each talk, you will have 3~4 minutes to answer the questions. Please write your answers on your answer sheet. Each answer will require just a word, phrase or short sentence. Please pay attention to grammar and spelling. Answers which cannot be understood will be considered incorrect. Please turn to the next page. [pause 4 seconds] Now we will begin Part III with questions no. 28 to 34.

For questions no. 28 to 34, listen to part of a lecture given at a conference. Before listening, you will have 30 seconds to read the instructions and questions no. 28 to 34. Please begin now. When you finish, look up from your desk. [Pause 30 seconds.]

Now listen to the lecture. You may take notes in your test booklet as you listen.

It is exciting to be a part of the medical field in this new century. There are so many innovations in hospitals today, innovations which you'll be able to encourage and direct. No longer do hospitals need to be the cold, impersonal institutions of the past. Then everything was kept clean, and efforts were made to make patients physically comfortable, but there was little attention to the mind, to the spirit. Now, of course, we're well aware of the effects of the mind and spirit on a person's physical health, and we know that a depressing hospital environment can actually slow down the healing process. So one of your concerns as administrators must be to create a hospital environment which "feeds" the patients' minds and spirits, so that they can get well sooner. In a little while, I'll be showing you some video footage of scenes at two hospitals where this principle is already being put into practice. For now, let me explain some of the innovations you'll be seeing.
One big change that is taking place in hospitals has to do with bringing patients in contact with nature. Studies have shown us that patients who can see trees and other greenery from their hospital windows actually recover faster than those who only see a brick wall or an air conditioner. There's something in human beings that makes us love to look at a tree. So as new hospitals are designed, an important consideration is whether patients' rooms have views of nature, and how much of the view can be seen while a patient is lying flat in bed. These changes are happening slowly, of course—I doubt that your hospital's board of directors would agree to tear down your old building and put up another one! But if you as administrators understand the principle, you will want to make what changes you can.

Research has shown that just a taste of nature can have significant healing effects. There are so many creative ways you'll be able to apply this principle. Even a painting of nature on a hospital wall, for example, can make a big difference to patients in intensive care units. One study in Sweden showed that patients who could see a nature painting had less anxiety and needed fewer strong pain drugs than patients with no painting. At one of the hospitals you'll be seeing on the video in a minute, the windows don't have a good view, but the curtains are printed with a colorful nature scene. At my own hospital, we let patients listen to CDs with sounds of birds singing and rivers bubbling over rocks. As I say, use your own imaginations to apply this principle of bringing nature into the healing process.

Other innovations in the healing process at hospitals don't require new buildings—only new attitudes. As I said, in the past, our number one concern for hospital rooms was that they be clean, free of things that might carry disease. But this attitude has to change in order to welcome some new partners into the healing process—animals. Many hospitals now have their own dogs and cats—gentle ones, of course, well-trained, kept clean—and patients love them. And this is the point—patients respond to them. They stop worrying about their own medical problems, their anxiety decreases, their blood pressure drops. As administrators, we need to stop thinking of animals as dirty. They may leave some stray hair around, but the benefits of animals to the life of a hospital can be enormous.

Now, let's go ahead and take a look at the video. (fading)
A few years ago, six thieves, pretending to be port officials, boarded a ship in Hong Kong, drugged the captain, and made off with $300,000 dollars worth of frozen food. The crew was lucky, as they survived the attack. Others have not been so lucky. **Twenty-three** crewmembers of a ship near the Philippines were killed when pirates attacked their ship.

Attacks occur worldwide, but there are a few areas where vessels are particularly at risk from pirates. Brazil, for example, is renowned for the high number of attacks on vessels in port. A reason for this could be the lack of security and police in Brazilian ports. For ships at sea, South East Asia is the place where crewmembers need to be the most alert. In this area, there are hundreds of little islands that pirates can hide on or escape to, which increases the likelihood of an attack on a ship.

If piracy is, in fact, an everyday occurrence, why don't we hear about it more on the news, or read about it in the newspapers? Well, the truth is that shipping companies hardly ever report attacks to the police after they happen. The companies are worried that if they report such incidences, the cost of insurance will increase and crewmembers will demand higher wages. Let's also not forget that the chances of recovering the cargo or finding the stolen ships again are quite slim. Ships are repainted, given new names and false papers. The cargo is sold or transferred to another ship. As for the pirates, they hardly ever get caught.

So, what measures are being taken to combat this problem? Most importantly, the International Maritime Bureau set up the Piracy Center in 1992. It's based in the Malaysian capital of Kuala Lumpur and financed by shipping and insurance companies. Its main purposes are to help ships report both attacks and attempted attacks, and to offer information and advice to shipping companies. The Center compiles detailed information about areas to avoid and measures to take to protect ships and crews.

In brief, not only is piracy a serious crime that has caused huge losses for shipping and insurance companies, but it is also a violent crime that has killed hundreds of innocent people. We need more support and cooperation from governments and organizations to put an end to a crime that should have been stopped centuries ago.

**Now you will have 4 minutes to answer questions no. 35 to 40. Write your answers on your answer sheet.**

[Pause 4 minutes.]  
This is the end of the Listening Comprehension Test.
Appendix 2: The GEPT Advanced Level Reading Pretest form

In this section, there are 20 questions altogether. You will have 50 minutes to complete this section. Please read the passages carefully to get the answers. Please write your answers on your answer sheet.
Part I

Read the following three articles and write answers to questions 1-13. Each answer will require just a phrase or short sentence.

Questions 1-5

U.S. oral surgeon Dr. Mansoor Madani says he has developed an alternative to tonsillectomies that works just as well as the traditional surgery, but has none of the drawbacks. This procedure, known as tonsillar coblation, uses radiofrequency-generated heat to shrink tonsils. In the procedure, Madani inserts a heat-emitting probe about the size of a needle into three spots on each tonsil, and keeps it in each position for around 10-15 seconds. The probe – heated to between 120 and 150 degrees Celsius – slowly shrinks the tonsils like a water balloon being stuck with a pin.

"It's just like when you put certain foods in the microwave and they shrink. It's very similar to microwave energy, but the heat isn't nearly as high," said Madani, an associate professor of oral and maxillofacial surgery at Temple University, and director of The Center for Corrective Surgery Techniques in Bryn Mawr.

Radiofrequency-generated heat has been used in the field of medicine for more than a century. It is currently being used in cardiac surgery, neurosurgery, orthopedic surgery, and urology. Madani himself had been using the technology in parts of the nose and mouth for several years, to treat chronic nasal congestion and snoring, before trying it out on tonsils.

Tonsils are the two lumps of tissue in the back of the throat on either side of the tongue, each about the size and shape of a large olive. Their purpose is to "catch" infections in the first two to three years of childhood, sampling the environment in order to help the child develop immunities, or antibodies. Yet in some cases, the tonsils also catch incoming germs, and this causes them to become infected. A medical study has shown that children who suffer from frequent tonsil infections (more than two or three times each year) are generally healthier after their tonsils are removed. And children who must have their tonsils removed have been found to suffer no loss in their resistance to infections.

Patients undergoing traditional surgery to have their tonsils removed usually need more than a week to recover, including one night in the hospital, and the remaining time at home. In addition, they may expect intense pain and severe bleeding. Many report subsisting for days on just ice cream and jello, as their throats are too sore to eat anything else.

By contrast, the tonsillar coblation is performed right in the doctor's office, with just a local anaesthetic or mild sedation. It is so quick that some choose to have it done on their lunch hour from work. Patients are discharged with a prescription for antibiotics, a special mouthwash, and a rinse. They can resume their normal activity immediately after the procedure, and most report suffering only mild discomfort.
Yet some skeptics warn that as with any new technology, unforeseen problems could surface farther down the road. "The procedure holds promise, but it is unclear whether the positive results will last," said Dr. Ralph Wetmore, chairman of the tonsils and adenoids subcommittee of the American Academy of Otolaryngology, and a professor at the University of Pennsylvania. "Certainly if you shrink the tonsil and it stays reduced, it would offer significantly less (of a chance of) morbidity than performing a tonsillectomy, but it's unknown what the long-term results are," Wetmore said.

Wetmore also pointed out that the tonsillar coblation may be difficult to perform on children because they would be awake. "Its benefits are probably better achieved in the adult population," he said.

1. What is the function of radiofrequency-generated heat in the procedure of tonsillar coblation?
   1. ____________________________________________________.

2-3. What are the two drawbacks of traditional tonsillectomies that the tonsillar coblation does not have?
   2. ____________________________________________________.
   3. ____________________________________________________.

4-5. What are the two potential problems of the tonsillar coblation as reported in the article?
   4. ____________________________________________________.
   5. ____________________________________________________.
Questions 6-9

Biotechnology is one of the big hopes among both politicians and stockholders. It is supposed to provide more jobs, promote more growth, assist the world's poor countries, create better harvests, and spare the Earth many pesticides in the bargain. But while genetic engineering produces new headlines almost daily, the raw material that this new science employs is slipping away. With mind-boggling speed, the plants and animals that genetic engineers depend upon are becoming extinct.

This disappearance affects all of us. It is true that no one knows exactly how many plants and animals there are on the planet. So far, 1.75 million species have been counted, and it is estimated there are far more. But before we can accurately say how many species exist, people are destroying them. As many as 130 species per day may be disappearing.

A commission of twelve professors assembled by the German government has sounded the alarm. The Scientific Council on the Global Environment has made a clear statement of its findings: The destruction of species due to human settlements and disturbance of landscapes and ecosystems has already reached such proportions that restoring a complex system of species might take millions of years if this "terrible trend" is not stopped within a few decades.

The commission points out that in its history to date, the Earth has seen five major extinctions. All of them were set off by natural catastrophes such as meteorite impacts or climate changes. The most recent of these waves of extinction occurred 65 million years ago. The present mass extinction, the sixth, is different, though, from all of its predecessors—not just in "the raging speed" at which it is proceeding but because, for the first time, it was unleashed by mankind. Homo sapiens is working toward a "total triumph" over millions of its fellow species, the 500-page document states. It may soon achieve "the most spectacular monument yet created by civilization"—an impoverished, monotonous world.

The detailed results are not yet available, but the consequences will be dramatic. In just a few more decades, the tropical rain forests could be completely destroyed through fire, logging, and man-made climate change. A similar fate may await the coral reefs, although here part of the blame will also be laid at the feet of amateur divers seeking pleasure underwater. And then there is the German tourist industry and its works in the Alps, with planned construction projects including 45 ski runs and the avalanche protective systems they will need, 22 reservoirs and hydroelectric projects, 49 highway projects, five power lines, and 18 resorts, according to the commission. Only when it has eaten up its natural capital, and the landscape can no longer pay dividends, will "the architects of the beautiful new Alpine world" realize what irreparable damage they have done there, the report says.

Just how thoroughly the 6 billion people on Earth have subjected it to their rule is obvious from a single statistic: Forty percent of all photosynthesis by green plants is already under human jurisdiction. The highest form of civilization apparently tolerates little competition for anything
other than its own crops. This is not just arrogant, it is dangerous. Only a few plant species, at risk from pests and climate changes, must provide food for more and more people. In the meantime, global feedback mechanisms, which are not yet understood, are being disturbed.

Although the economy is usually blamed when suspects are sought for this terrible situation, economic theory provides no justification for it. On the contrary, even ostensibly useless or undiscovered species have potential value. The loss of Earth's biological diversity cannot be quantified. When preservation interests come into competition with development, they almost always lose—even under an environmentally friendly regime. But now Chancellor Gerhard Schroder and his colleagues must at least allow their commission to be heard. Its conclusion is that something must be done if the earth's biological diversity is to be preserved.

6. What has the German government done that shows concern for the environment?

7. How is the sixth extinction unlike earlier ones?

8. What two physical features does the report predict may soon disappear from the landscape?

9. What is the tone of this article?
   A. Complimentary
   B. Indifferent
   C. Critical
   D. Encouraging
The great Nigerian novelist Chinua Achebe, one of the founders of African literature, celebrates his 70th birthday this month. Out of his multifaceted work, the novels of his African Trilogy, written between 1958 and 1966, especially stand out. The first, *Things Fall Apart*, has sold almost 10 million copies worldwide and has become Africa's best-known novel.

A decisive motivation for Achebe's writing was his dissatisfaction with the colonial image of Africa presented by Europeans in such influential books as Joseph Conrad's *Heart of Darkness*. Edgar Wallace's "Sanders" books depicted this view of Africa with especially dismal directness. They were "books in which blacks stank, danced, screamed, lied, and begged," as one reviewer has summarized them.

Achebe rejected the racist arrogance of such depictions, as well as their tendency to make Africans vehicles for psychological, moral, or political lessons. What was missing was any genuine interest in Africans, their art, religion, or history, or an ear open to hearing their own stories.

In order to present the real Africa, Achebe needed only to recall his own childhood. His father worked for the Christian Missionary Society. In the eastern Nigerian village where the writer was born in 1930, other members of his family continued to worship Chukwu, the god of creation, and Ani, the Earth goddess. The young Achebe moved within both worlds.

*Things Fall Apart*, the first volume of the African Trilogy, set in the 1890s, tells about the village of Umuofia and one of its most powerful men, Okonkwo. He is the son of a weakling who was never able to attain the honorary titles of a real man. While the son does become wealthy through hard work, he is plagued by fears of failure, which express themselves through stuttering and violent outbursts. When the British colonizers and missionaries arrive in Umuofia and attempt to impose the Christian religion and a new social order, Okonkwo reacts violently and strikes one of the court officials dead during a meeting of clan members. To avoid his punishment, Okonkwo hangs himself, a violation of the clan's traditions. The old order is shattered. It cannot withstand the challenge of the new, but like the protagonist, it also fails because of its inner weaknesses.

In 1960, Achebe published *No Longer at Ease*, with Okonkwo's grandson Obi as the main character. With the financial support of his village, Obi had gone to study in England, and upon his return to Nigeria had managed to get a minor bureaucratic job. His clan members expected favors and reimbursement of his stipend so that other young men from the village could go to school.

But Obi went off instead to Lagos and fell into the corruption of urban life. He embezzled and then, like his grandfather, was punished by the white justice. So this wanderer between two worlds became a victim of the tribalism and corruption that continues to be responsible for many of Africa's failings.

The third novel of the trilogy, perhaps the most beautiful and best written, appeared in 1964. *Arrow of God* takes place in the 1920s, and Ezeulu, a priest of the Ibo religion, is the central character. Once again, the conflict is between Ibos and the English, tradition and modernity. The
British colonial officials impose a policy to turn over local government to responsible local people. They choose Ezeulu because of his good reputation. However, he refuses the honor and is imprisoned.

As a result, he cannot preside over the festival of the new moon. This means that the yam harvest is not allowed to begin, and the villagers begin to starve. Some turn to the new Christian church for a faith which will allow them to eat the yams.

In this book, Achebe shows how the opposing parties in Umuaro and the hierarchical British authorities are concerned about the same values of rank and respect. Achebe presents each voice so authentically that the reader is able to understand that knowledge and authority are not "naturally" to be found on one side or another. Achebe has restored to Africans their voices. In the trilogy and his later works, he writes in an English that is permeated with the structures and rhythms of the Ibo language, thus linking his new stories to traditional myths, fables, and proverbs.

In Achebe, we hear the most powerful yet elegant voice of the second half of the 20th century in Africa, a tireless teacher and moralist and, above all, a powerful storyteller.

10. What specifically prompted Achebe to begin writing novels?

11. In *Things Fall Apart*, what do the main character and the traditional order have in common?

12. How does Achebe present Europeans in his novels?
   A. As superior to blacks in most ways
   B. As inferior to blacks in most ways
   C. As gentle but ambitious
   D. As benevolent but interfering

13. In the novels discussed in this article, what is distinctive about Achebe's writing style?
Questions 14-20

The Perfume Hunters
Sniffing out new smells for use in cosmetics and household products involves blood, sweat and plenty of insect repellent.

Tired, scratched and soaked in sweat, the hunters begin to think of turning back. Time is running out. Dusk is falling and they still haven't caught sight of their quarry. Suddenly they stop. One of the men lifts his head and sniffs. He knows they are close. He scans the undergrowth in the deepening gloom—and suddenly he spots what they have been looking for. There, hidden beneath some leaves at nose height is a tiny spike of flowers, the whole bunch no bigger than a thumbnail. Within minutes, the hunters have set their trap. All they have to do now is wait.

The hard work was worth it. The next morning, there in the trap is a rare catch—a new sort of smell. For the men in the Madagascan forest are perfume hunters. And instead of rifles, they are armed with nothing more sinister than a few glass jars, a couple of pumps and some tubing which they will use to capture new and exciting fragrances to make our lives smell sweeter.

Ever since the *unguentari* pldied their trade in ancient Rome, perfumers have had to keep abreast of changing fashions. These days they have several thousand ingredients to choose from when creating new scents, but there is always a demand for new combinations. The bigger the "palette" of smells, the better the perfumer's chance of creating something new and fashionable. Even with everyday products such as shampoo and soap, consumers are becoming increasingly fussy. Cheap, synthetic smells are out. Fresh, natural smells are in. And many of today's fragrances have to survive tougher treatment than ever before, resisting the destructive power of bleach or a high temperature wash cycle.

Chemists can now create new smells from synthetic molecules, but nature has been in the business far longer. Plants produce countless fragrant chemicals. Many are intended to attract pollinators. Others are produced for quite different purposes. The fragrant resins that ooze from wounds in a tree, for example, defend it against infection.

The island of Madagascar is an evolutionary hot spot; 85% of its plants are unique, making it an ideal source for novel fragrances. So last October an expedition, including Robin Clery, a chemist, and Claude Dir, a perfume company director, explored two contrasting landscapes in northern Madagascar. Their first stop was a remnant of rainforest in the national park of Montaigne d'Ambre. The second was the tiny uninhabited island of Nosy Hara off the northwest coast.

With some simple technology, borrowed from the pollution monitoring industry, and a fair amount of ingenuity, the perfume hunters bagged 20 promising new aromas in the Madagascan rainforest. Each day the team set out from their "hotel"—a wooden hut lit by kerosene lamps, and trailed up and down paths and animal tracks, exploring the thick vegetation up to 10 meters on either side of the trail. Some smells came from obvious places, often big showy flowers within easy reach. Others were harder to pin down. "Often it was the very small flowers that were much more interesting," says Clery.
In fact, some of the most promising fragrances were given off by resins that oozed from the bark of trees. Resins are the source of many traditional perfumes such as frankincense and myrrh. The most exciting resin that the team found came from a *Calophyllum* tree, a relative of the Asian beauty leaf, which produces a strongly scented medicinal oil. The sap smelt rich and aromatic, but it also smelt of something the fragrance industry has had to learn to live without—*castoreum*, a substance extracted from the musk glands of beavers and once a key ingredient in many perfumes. "We don't use animal products any longer," says Dir, "so to find a tree with an animal smell is extremely precious."

After the luxuriance of the rainforest, the little-known island of Nosy Hara was a stark, dry place—geologically and biologically very different from the mainland. "Apart from two beaches, the rest of the island is impenetrable, except by hacking through the bush," says Clery. One of the biggest prizes here was a sweet-smelling sap weeping from the gnarled branches of some ancient shrubby trees in the parched interior. So far no one has been able to identify the plant.

The group also set out from the island to capture the smell of coral reefs. Odors that conjure up sun-kissed seas are highly sought after by the perfume industry. "From the ocean, the only thing we have is seaweed, and that has a dark and heavy aroma. We hope to find something unique among the corals," says Dir.

The challenge for the hunters was to extract a smell from water rather than air. This was an opportunity to try Clery's new "aquaspace" apparatus—a set of filters that work underwater. On Nosy Hara, jars were fixed over knobs of coral about 2 meters down and water pumped out over the absorbent filters. So what does coral smell like? "It's a bit like lobster and crab," says Clery.

The team's task now is to recreate the best of their captured smells. First they must identify the molecules that make up each fragrance. Some ingredients may be quite common chemicals. But some may be completely novel, or they may be too complex or expensive to make in the lab. The challenge then is to conjure up the fragrances with more readily available materials. "We can avoid the need to import plants from the rainforest by creating the smell with a different set of chemicals from those in the original material," says Clery. "If we get it right, you can sniff the sample and it will transport you straight back to the moment you smelt it in the rainforest."

**Summary**

The perfume trade has a long history, dating back to (14)_______________. Today, perfumers can choose from a wider range of chemicals, and many of these are synthetic. However, fresh, natural fragrances are more (15)______________ and perfumers continue to hunt for new smells from nature. Plants are a major source of perfumes, producing smells for many reasons, such as to encourage useful insects and to prevent (16)_______________. Last October, perfume hunters traveled to Madagascar, a promising site for new smells because of the (17)______________ of its trees and flowers. In a rainforest, the group collected an extremely valuable smell which resembled a chemical called *castoreum*. This new smell was considered very useful because today perfumers have stopped employing (18)_______________. At a small island, the fragrances of (19)_______________ were collected by the same team using equipment that works underwater. On returning to the laboratory, the group will attempt to reproduce the new smells using chemicals that are (20)_______________.

-The End-
ADVANCED READING COMPREHENSION TEST

SECTION 2: SKIMMING & SCANNING

In this section, there are 20 questions altogether. You will have 20 minutes to complete this section. Please read the passages quickly and selectively to get the answers. Please write your answers on your answer sheet.
Part I

In this part, there are 3 passages with a total of 6 multiple-choice questions. For each passage, please first read the question(s) and then read the passage quickly and selectively to find the answer(s). Please write your answers on your answer sheet.

Question 21

First read the following question.

21. What is the main purpose of this passage?
   A. To announce a forthcoming activity
   B. To demonstrate computer software
   C. To explain an office regulation
   D. To introduce new technology

Now skim the following text to find the answer to the above question.

Even with computerized information storage systems, modern offices continue to generate a lot of used paper. In Japan, for example, paper accounts for around 40 percent of office waste. Until now the most that environmentally-conscious companies have been able to do about the one-way paper flow is to ask their staff to write fewer memos and to use both sides of each sheet of paper before consigning it to the recycling bin. It has not been possible for office workers to reuse the paper itself because most inks contain a dye known as carbon black, which is very difficult to remove, or "decolor," from paper.

Recently, though, Toshiba has developed a new kind of ink, one which does not contain carbon black but does contain a decoloring agent mixed with the dye. When subjected to temperatures above 356 degrees Fahrenheit, the chemical bonding of the ingredients in the ink breaks down, and the decoloring agent erases the ink from the page. By using the new ink, offices could easily do their own paper recycling.
Question 22

First read the following question.

22. According to the following passage, what experience have the winners of the 15th Future Magazine Innovation Awards had in common?
   A. They have suffered unexpected setbacks.
   B. They have saved the lives of a lot of people.
   C. They have been the leaders of university research teams.
   D. They have achieved breakthroughs in surprising circumstances.

Now skim the following passage to find the answer to the above question.

**The 15th Future Magazine Innovation Awards**

Even the most brilliant scientists tend to be caught off guard when inspiration strikes. The nine winners of our 15th Future Magazine Innovation Awards all report that the best ideas come to mind when—and where—you'd least expect them. One of our winners thought up a concept that may revolutionize telecommunications as he was retrieving his car from a university parking lot. Another hatched a plan to save millions of people from being blown up by land mines as he and a colleague ate lunch at an outdoor restaurant. Still another was rendered speechless when a way to make plastic from corn came to him out of the blue as he walked from his desk to his lab bench.

As we hurtle headfirst into the new millennium, it can't hurt to remember that the world can change because one human mind suddenly and unexpectedly cast up a gem of an idea. One person's imagining benefits us all.
Questions 23-24

First read the following questions.

23. What is the main subject of this passage?
   A. The style of a renowned glass artist
   B. The revival of glass architecture
   C. The uses of glass in theater design
   D. The glass buildings constructed in the 1950s

24. What is the purpose of the third paragraph?
   A. To explain the reasons for a phenomenon
   B. To compare two different viewpoints
   C. To point out some potential problems
   D. To describe a production process

Now skim the following passage to find the answers to the above questions.

In the 1920s, the renowned Swiss architect Le Corbusier observed that the history of architecture is the history of the window. In the new generation of glass buildings now appearing in cities around the globe, "the window has broken out of the frame," says architect James Polshek, lead designer of the Rose Center for Earth and Space at the American Museum of Natural History in Manhattan.

From Berlin to Tokyo, New York to London, a rash of striking new structures is giving glass construction new life. Not since the early 1950s, when sleek, green-tinted glass buildings like New York City's Lever House rose amid the stone canyons of countless major cities, has glass elicited so much attention. There are so many glass buildings currently on the drawing boards or under construction, in fact, that it is hard to keep track of them all. From the 357-foot-long barrel-vaulted, transparent roof of New York-based architect Rafael Vinoly's Kimmel Center for the Performing Arts in Philadelphia to the soaring glass dome of London architect Norman Foster's restoration of the historic German Reichstag, glass is assuming innovative, dramatic and surprising forms.

This new trend is due primarily to the development of new construction methods and glass production techniques. Ingenious systems now allow expanses of glass to be suspended in front of a series of nearly invisible metal trusses and braces. In addition, advancements in glass production since the 1973 energy crisis have resulted in revolutionary changes in its performance. Glass is now being manufactured with special coatings that adjust the thermal intake for all four directions of the compass and even allow for the increased intensity of the sun at certain latitudes. The new glass is not only spectacular to look at and through, it is also safer, stronger and more energy efficient.
Questions 25-26

First read the following questions.

25. Who is most probably the writer of the passage?
   A. The spokesperson for a human rights organization
   B. A former employee of the Prison Bureau
   C. The chief of the Prison Bureau
   D. A prisoner in an overcrowded prison

26. The passage mentions a recent change in funding for prisons. What is the writer’s attitude toward this change?
   A. It seems promising.
   B. It will cause more problems.
   C. It is a disappointing decision.
   D. It will have both positive and negative effects.

Now skim the following passage to find the answers to the above questions.

**OPINION**

**Locking up criminals is an easy option**

For the past five and a half years, I have been calling for increased investment in better treatment of and conditions for prisoners, aimed at protecting the public by preventing reoffending, and so I am delighted that the government is planning to give additional money to the Prison Bureau. The dirty and overcrowded state of our prisons, and the current rate of reoffending, are both by-products of the lack of consistent and well-planned investment, not just in prisons but in society as a whole, improvements to both of which now appear possible.

I took over at a time when the morale in the Prison Bureau was at rock bottom. Reports attacking the Bureau over two escapes by prisoners had been followed by the resignation of the former chief. The prison population was rising rapidly, by more than 30 percent, while, at the same time, a 13.5 percent reduction was being imposed on its budget.

The politicians’ cry that “prison works,” echoed by some members of the public, encouraged this increase in the prison population. This is a cry that is true to the extent that people in prison cannot commit crimes outside. But it ignores the fact that all except a very few of those in prison are going to come out. These people will be more likely to reoffend if nothing is done, in prison, to tackle their offending behavior. Therefore, I sincerely expect that, with the new government funding, more resources can be allocated to the provision of full, purposeful and active days for all types of prisoners, designed to challenge offending behavior and enable them to live useful and law-abiding lives on release.
About a millennium ago, legends tell us, a Viking named Leif Eriksson sailed to the shores of North America, arriving hundreds of years ahead of Christopher Columbus. Even though archaeologists have yet to uncover any physical evidence of Eriksson's visit, the presumption that a Viking band traveled that far has gained credibility in recent years. Excavations in Greenland indicate that Vikings flourished there for hundreds of years, trading with the European continent and probably Native American tribes, before disappearing.

27. ______
A central figure in this story was Eriksson's father, Erik the Red, who grew up in Iceland. In 980 A.D., Erik the Red headed farther west when he was banished from Iceland--for murder. He set sail for land that was visible west of Iceland. Three years later, he returned to Iceland and convinced hundreds of others to join him in settling this new country. Some 25 boats set out for what Erik the Red had dubbed Greenland. Only 14 ships survived the seas, but about 450 new colonists set foot ashore.

28. ______
The land they saw before them was bare, uninhabited, and inhospitable, but Erik the Red's advertisements were not entirely false. A thin green carpet of arctic heath promised support for grazing farm animals. Farms sprang up quickly and, later, churches. One colony, simply called the Eastern Settlement, sat in the toe of Greenland; the Western Settlement lay close to what is now Nuuk, Greenland's capital.
Settlement a Challenge

Settling Greenland posed a formidable challenge. There were no trees large enough to produce timber for shelter or fuel. The only wood was small brush and driftwood. The Vikings settled inland, on fjords resembling those of their homeland. There they built homes of driftwood, stone, and sod. For adequate insulation, the walls of some buildings were made six to 10 feet thick.

29. _______

Shelter, food, and clothing were, of course, essential to survival. The summer was too short to farm grain crops, so settlers probably went without beer or bread. Although they farmed domesticated animals imported from Europe--goats, sheep, cattle--the settlers ate them sparingly, relying instead on secondary products, such as milk and cheese. In the early days, the Greenlanders' lives differed little from those of their compatriots in Scandinavia. They netted fish and hunted seal and caribou. They wove clothing from wool and linen, sometimes adding the fur of the arctic hare.

Trade with Scandinavia

For about two centuries, Greenland's Vikings had the country to themselves. Yet life was by no means easy, and they relied on a fragile trade with Scandinavia to survive. In exchange for iron, timber, and grain from Europe, they traded pelts of bear and arctic fox as well as narwhale tusks and rope made of walrus hide. Whalebone, too, was traded to Europeans for use in stiffening clothes. According to one account, the Greenlanders even traded live polar bears.

30. _______

At some point during the fourteenth century, Greenland's climate grew colder. With the climate change, glaciers began creeping over the land, bringing with them a runoff of sand, silt, and gravel. That runoff slowly robbed the settlers of valuable pastureland. To make matters worse, the Black Death hit Iceland, killing some 30 percent. Although there is still no evidence the sickness reached Greenland, archaeologists believe it left its mark by curtailing the flourishing trade.

31. _______

The Greenlanders adapted. Recent evidence shows that their diet shifted from land-based foods to marine products. Like their kin in Norway, the Vikings in Greenland had always exploited marine life but, by the close of the fourteenth century, the proportion of their food taken from the sea had risen to 80 percent.

32. _______

Between 1100 and 1200 A.D., as the colder weather arrived, so did the Thules. These Native Americans, migrants from the area surrounding the Bering Strait began trickling eastward from Ellesmere Island, just northwest of Greenland. It's likely that an uneasy trade between the Vikings and Thules sprang up and that, as living conditions grew harsher for the Vikings, the better-adapted Thules thrived.

Greenland Settlements Abandoned

The Western Settlement was abandoned by 1350 A.D. and the Eastern Settlement by 1500 A.D. When asked what became of the Vikings, Danish archaeologist Jette Arneborg says she thinks they struggled mightily to adapt to the increasingly difficult conditions. But as the weather worsened and life became even harsher, some may have returned to Iceland. And it's easy to imagine that, as trade dwindled, the settlements may have become so depopulated the colonists simply were unable to replace themselves.
Part III

In this part, there are 3 passages with 12 questions. Please first read the questions and then read the passages quickly and selectively to find the answers. For each question, mark A, B, or C on your answer sheet.

Questions 33-40

First read the following questions.

(A=Middlebury; B=Bowdoin; C=Colgate)

_____ 33. Which school is located close to a large city?
_____ 34. Which school has the largest percentage of minority students?
_____ 35. Which school has announced plans to hire more teachers?
_____ 36. Which school recently built a new cultural center?
_____ 37. Which school did writer Hawthorne graduate from?
_____ 38. Which school offers students summer research opportunities?
_____ 39. Which school requires students to take a writing course?
_____ 40. Which school claims that its strongest programs are in the sciences?

Now scan the following passages to find the answers to the above questions.

Middlebury College

Middlebury College, located in Middlebury, Vermont, is one of New England's leading small, residential liberal arts colleges. It offers students a broad curriculum embracing the arts, humanities, literature, foreign languages, social sciences, and natural sciences. In addition, the College's 350-acre campus has been said to be "among the prettiest in the world."

Middlebury believes that the purpose of the liberal arts curriculum is to give every student a detailed knowledge of at least one subject, and to correlate this with a broad understanding of the liberal arts. To achieve this objective, students are required to work intensively in one or more departments, while also completing electives in fields outside of their specialization. All students must complete a major, a first-year seminar, a college writing course, and two units of physical education.

Middlebury's undergraduate program is greatly enriched by its other programs. Every summer, the main campus is devoted completely to the study of eight foreign languages and cultures. At the same time, at the nearby Bread Loaf campus, the Bread Loaf School of English is in session.

Currently, there are 2,265 students enrolled at Middlebury, of whom 12% are members of minority groups, and 95% are from out of state. By the year 2005, the College plans to increase the size of its student body to around 2,350. Middlebury's full-time faculty of 218 is also expected to increase to nearly 250 by that time, enabling the College to further deepen and strengthen its academic programs.

And to better serve this enlarged community, Middlebury is in the process of constructing major new facilities. In addition to a new hockey rink and a new science center, planned capital projects include an expansion and renovation of Starr Library, new dining facilities, expanded student activities space, new student residence halls, and a new humanities center.
Bowdoin College

Bowdoin College's beautiful 110-acre campus is located in Brunswick, Maine. Brunswick, one of New England's most attractive college towns, is just 42 kilometers from Maine's largest city of Portland, and a two-hour drive from Boston.

The alma mater of literary giants Nathaniel Hawthorne and Henry Wadsworth Longfellow, Bowdoin has undergone significant changes in recent years. The College now boasts two new, state-of-the-art science facilities and new residential halls. It has gradually phased out its fraternities and sororities, and instituted a new College House system designed to promote interaction among diverse groups of students.

Bowdoin currently enrolls 1,608 students, of whom 13% are members of minority groups, and 82% come from out of state. The school has a full-time faculty of 113, for a student/faculty ratio of 10:1.

Bowdoin's general distribution requirements ensure that all graduates gain a strong foundation in humanities and the arts, natural sciences and math, social sciences, and non-Eurocentric studies. At the same time, Bowdoin is unusual among liberal arts colleges because its strongest programs are in the sciences. Self-designed and double majors have become increasingly popular among Bowdoin undergraduates, and about 80 percent of juniors and seniors conduct independent study programs with faculty members. Students can also elect unusual research opportunities, such as participation in Arctic archaeological research in Labrador or ecological research in the Bay of Fundy, Canada.

During the nonacademic portion of the year, Bowdoin opens its doors to people from all walks of life. Bowdoin College Summer Programs consist of educational seminars, professional conferences, sports clinics, specialized workshops, and occasional social events, and they attract several thousand people to the College each summer.

Colgate University

Colgate University is located in the village of Hamilton, at the northern end of the Chenango Valley, in upstate New York. Its 515 acres of campus begin at the village edge on the valley floor and rise to a forested hill.

Colgate currently enrolls 2,866 students, of whom 68% come from outside of New York State, and 14% are members of minority groups. Its full-time faculty of 230 gives Colgate a student/faculty ratio of 11:1.

Colgate is currently in the process of expanding and renovating its campus. Case Library was recently renovated, new housing has been built, and a social sciences academic building, cultural center, and fitness center have been added. Residence hall renovation continues, and a new academic facility for the arts is under construction.

Colgate offers 50 undergraduate concentrations (majors), in four academic divisions: Humanities, Natural Sciences and Mathematics, Social Sciences, and University Studies. Competency must be demonstrated in a foreign or classical language, and in English composition. First-year students enroll in a first-year seminar during the fall term.

In addition, Colgate offers a small graduate program leading to the Master of Arts (M.A.) in several academic fields, and the Master of Arts in Teaching (M.A.T.) degree.

During the summer, Colgate encourages its students to take part in one of the many research projects being carried out on campus. Each year, more than 100 Colgate undergraduates receive summer research assistantships, enabling them to work full-time on research or scholarly projects in close collaboration with faculty members. Typical research appointments are for eight- to ten-week periods. During this time, partially subsidized on-campus housing is available, and special academic and recreational events enhance this scholarly community at Colgate.

-The End-
LR-64
### Appendix 3: The GEPT Advanced Level Listening and Reading answer keys

#### 1. Listening Comprehension Test

**Part I: Short Conversations & Talks**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>6</td>
<td>A</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>7</td>
<td>C</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>8</td>
<td>B</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>9</td>
<td>A</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>10</td>
<td>C</td>
<td>15</td>
</tr>
</tbody>
</table>

**Part II: Long Conversations**

<table>
<thead>
<tr>
<th>Questions 16-20</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. (at) <strong>university</strong></td>
</tr>
<tr>
<td>17. <strong>after</strong></td>
</tr>
<tr>
<td>18. <strong>psychological thriller(s)</strong></td>
</tr>
<tr>
<td>19. (a mixture of) <strong>fear and curiosity</strong> (about ourselves)</td>
</tr>
<tr>
<td>20. C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions 21-27</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. <strong>migration</strong> or <strong>migrants</strong></td>
</tr>
<tr>
<td>22. <strong>migrant workers</strong></td>
</tr>
<tr>
<td>23. <strong>internal displacement</strong> or people forced to <strong>move</strong> to somewhere else <strong>within their own countries</strong></td>
</tr>
<tr>
<td>24. (are forced to) <strong>abandon</strong> their <strong>land</strong></td>
</tr>
<tr>
<td>25. the <strong>environment</strong> has been <strong>destroyed</strong></td>
</tr>
<tr>
<td>26. <strong>rural areas</strong></td>
</tr>
<tr>
<td>27. <strong>urban areas</strong></td>
</tr>
</tbody>
</table>

**Part III: Long Talks**

<table>
<thead>
<tr>
<th>Questions 28-34</th>
</tr>
</thead>
<tbody>
<tr>
<td>28. <strong>mind</strong> or <strong>spirit</strong></td>
</tr>
<tr>
<td>29. (bringing patients in) <strong>contact with nature</strong></td>
</tr>
<tr>
<td>30. (arranging patient) <strong>rooms with views of nature</strong></td>
</tr>
<tr>
<td>31. (hanging) <strong>paintings of nature</strong> (on hospital walls)</td>
</tr>
<tr>
<td>32. (choosing/designing) <strong>curtains with</strong> (colorful) <strong>nature scenes</strong></td>
</tr>
<tr>
<td>33. (letting patients listen to CDs with) <strong>sounds of nature</strong>/birds and rivers</td>
</tr>
<tr>
<td>34. (bringing) <strong>animals</strong> (into the healing process) or <strong>new attitudes</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions 35-40</th>
</tr>
</thead>
<tbody>
<tr>
<td>35. <strong>increasing</strong></td>
</tr>
<tr>
<td>36. (more and more) <strong>violent</strong></td>
</tr>
<tr>
<td>37. <strong>security</strong> is <strong>poor</strong> or lack of police</td>
</tr>
<tr>
<td>38. (there are) <strong>many islands</strong> where <strong>pirates</strong> can <strong>hide</strong></td>
</tr>
<tr>
<td>39. <strong>insurance cost and wages</strong></td>
</tr>
<tr>
<td>40. to offer <strong>information and advice</strong> (to shipping companies)</td>
</tr>
</tbody>
</table>
2. Reading Comprehension Test

Part I: Careful Reading

Questions 1-5

1. to shrink tonsils
2. They usually need more than a week to recover.
3. They may expect/experience intense pain (in the throat) and severe bleeding.
4. Positive results may not last. or It's unknown what the long-term results are.
5. It may be difficult to perform on children.

Questions 6-9

6. It has assembled a commission (of twelve professors) to study human effects on the environment.
7. It is of human origin. or It is proceeding at a raging speed.
8. tropical rainforest and coral reefs
9. C

Questions 10-13

10. his dissatisfaction with the colonial image of Africa in earlier novels about Africa
11. They both fail because of inner weakness.
12. D
13. He writes in an English that is permeated with the structures and rhythms of the Ibo language.

Questions 14-20

14. ancient Rome
15. popular
16. infection
17. unique nature
18. animal products or animals or beavers
19. coral(s) or coral reef(s)
20. readily available or produced in a lab

Part II: Skimming & Scanning

<table>
<thead>
<tr>
<th>Questions 21-26</th>
<th>Questions 27-32</th>
<th>Questions 33-40</th>
</tr>
</thead>
<tbody>
<tr>
<td>25. C</td>
<td>31. F</td>
<td></td>
</tr>
<tr>
<td>26. A</td>
<td>32. G</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4: Results of the student questionnaire

1. The listening test

3. The test instructions in this listening test are:

4. The speech rate in this listening test is:

5. The questions in this listening test are:

6. To me, the amount of time provided for responding to Part I in this listening test is:

7. To me, the questions in Part I are:

8. To me, the amount of time provided for responding to Part II and III in this listening test is:

9. To me, the amount of time provided for reading the questions in advance of each conversation/talk in Part II and III is:

10. To me, the questions in Part II are:

LR-67
11. To me, the questions in Part III are:

- Easy: 3%
- Just right: 38%
- Difficult: 44%
- Very difficult: 15%

12. In my opinion, this listening test can reflect my listening ability.

- Agree: 49%
- Barely agree: 43%
- Disagree: 9%

13. I expect to achieve _____ percentage of correct items in this listening test.

- Above 80%: 11%
- 60-80%: 48%
- Below 60%: 41%

2. The reading test

14. The test instructions in this reading test are:

- Clear: 63%
- Acceptable: 29%
- Unclear: 9%

15. The questions in Part I (Careful Reading) are:

- (number of items)
  - Slightly few: 3%
  - Just right: 58%
  - Slightly many: 38%
  - No answer: 1%

16. The amount of time provided for responding to the questions in Part I (Careful Reading) is:

- Slightly short: 49%
- Just right: 45%
- Slightly long: 6%

17. To me, the questions in Part I (Careful Reading) are:

- (difficulty level)
  - Easy: 4%
  - Just right: 47%
  - Difficult: 40%
  - Very difficult: 9%
18. In my opinion, Part I (Careful Reading) can reflect my reading ability.

<table>
<thead>
<tr>
<th>Agree</th>
<th>39%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barely agree</td>
<td>11%</td>
</tr>
<tr>
<td>Disagree</td>
<td>1%</td>
</tr>
<tr>
<td>No answer</td>
<td>0%</td>
</tr>
</tbody>
</table>

19. I expect to get _____ of answers right in Part I (Careful Reading) of this reading test.

<table>
<thead>
<tr>
<th>Above 80%</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60-80%</td>
<td>46%</td>
</tr>
<tr>
<td>Below 60%</td>
<td>37%</td>
</tr>
<tr>
<td>No answer</td>
<td>2%</td>
</tr>
</tbody>
</table>

20. The questions in Part II (Skimming & Scanning) are: (number of items)

<table>
<thead>
<tr>
<th>Slightly few</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just right</td>
<td>22%</td>
</tr>
<tr>
<td>Slightly many</td>
<td>75%</td>
</tr>
<tr>
<td>No answer</td>
<td>2%</td>
</tr>
</tbody>
</table>

21. The amount of time provided for responding to the questions in Part II (Skimming & Scanning) is:

<table>
<thead>
<tr>
<th>Slightly short</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just right</td>
<td>16%</td>
</tr>
<tr>
<td>Slightly long</td>
<td>3%</td>
</tr>
</tbody>
</table>

22. To me, the questions in Part II (Skimming & Scanning) are: (difficulty level)

<table>
<thead>
<tr>
<th>Easy</th>
<th>3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just right</td>
<td>45%</td>
</tr>
<tr>
<td>Slightly difficult</td>
<td>39%</td>
</tr>
<tr>
<td>Very difficult</td>
<td>13%</td>
</tr>
</tbody>
</table>

23. In my opinion, Part II (Skimming & Scanning) can reflect my reading ability.

<table>
<thead>
<tr>
<th>Agree</th>
<th>27%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barely agree</td>
<td>44%</td>
</tr>
<tr>
<td>Disagree</td>
<td>27%</td>
</tr>
<tr>
<td>No answer</td>
<td>2%</td>
</tr>
</tbody>
</table>

24. I expect to get _____ of answers right in Part II (Skimmin & Scanning) of this reading test.

<table>
<thead>
<tr>
<th>Above 80%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60-80%</td>
<td>35%</td>
</tr>
<tr>
<td>Below 60%</td>
<td>54%</td>
</tr>
<tr>
<td>No answer</td>
<td>6%</td>
</tr>
</tbody>
</table>
General English Proficiency Test
Advanced Level Writing Test

Contents

I. Pretest Development Process............................................................... W-2
II. Rating Process and Scoring Methods............................................... W-4
III. Descriptive Analysis of the Pretest Results.................................. W-5
IV. Conclusion and Suggestions....................................................... W-11
V. Appendices..................................................................................... W-13
   Appendix 1: The GEPT Advanced Level Writing Pretest form........ W-13
   Appendix 2: Rating scales for the GEPT Advanced Level Writing Test.. W-18
   Appendix 3: Results of the questionnaire survey.............................. W-19
   Appendix 4: Advanced level writing samples.................................. W-22
I. Pretest Development Process

The GEPT Advanced Level Writing Test aims at assessing the English writing ability of English majors graduating from Taiwan's universities or colleges or those who had studied and obtained certificates of graduation in undergraduate or postgraduate programs in English-speaking countries. Those who pass this test should be able to write English abstracts, reports, theses or dissertations, and news articles, and express themselves clearly in English on a wide variety of topics.

1. Trial test

Following the above results, the GEPT team produced a prototype for the advanced level writing test in 2000. After several meetings, the GEPT committee members decided in April 2001 that a trial test be conducted on four item types, including abstracts, papers, letters, and reports. The design of a trial form and a questionnaire was completed by the research team in May and reviewed and approved by the committee members in June before a two-stage trial test was conducted in order to determine

- whether the specifications for the four item types were appropriate;
- whether examinees' writing ability could be effectively elicited;
- whether the test duration was appropriate;
- whether the rating scales were appropriate;
- whether difficulties arose in the rating process.

After the trial test, the research team made necessary modifications to the item type, difficulty level, rating scales, and test method of the advanced level writing test on the basis of the trial test results, interviews with the examinees, and observations of the test administrators, and had the revised version approved by the research committee in September 2001.

The advanced level writing test consisted of two tasks, and the duration of the test was 105 minutes. Task 1: Examinees had to read two interrelated verbal inputs in English and write an essay of 250 words in which examinees had to summarize the main points of the two verbal inputs and give personal opinions. Each of the verbal inputs contained approximately 400 words and was of the GEPT High-Intermediate difficulty level. The amount of time provided for carrying out Task 1 was sixty minutes. Task 2: Examinees had to interpret two interrelated non-verbal inputs and write under the given instructions an output of 250 words in which they summarized the inputs, discussed possible causes, and made recommendations. The amount of time provided for carrying out Task 2 was forty-five minutes (see Table 1).

Table 1  Item type of advanced writing

<table>
<thead>
<tr>
<th>Task</th>
<th>Input</th>
<th>Output</th>
<th>Length</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Verbal input: 2 texts</td>
<td>summary + expressing opinions</td>
<td>250 words</td>
<td>60 min.</td>
</tr>
<tr>
<td>Task 2</td>
<td>Non-verbal input: 2 charts, graphs, tables, or pictures</td>
<td>summary + discussing possible causes + making recommendations</td>
<td>250 words</td>
<td>45 min.</td>
</tr>
</tbody>
</table>
A five-band analytical scoring was used for rating the advanced level writing test performance rather than the holistic scoring adopted for use in the other levels of the GEPT writing tests. The reason why analytical scoring was used was that the number of advanced level examinees was quite small, and also holistic scoring might not effectively discriminate well between strong examinees' writing ability. In addition, the results of the two trial tests showed that examinees' writing abilities tended to develop idiosyncratically. For example, an examinee might perform differently in English structures, sentence patterns, wording, and summaries, and holistic scoring seemed unable to profile the examinee's writing ability. The revised analytical scoring used in the advanced level writing test consisted of four parts: relevance and adequacy, coherence and organization, lexical use, and grammatical use, and was aimed at assessing the range, accuracy, and appropriacy of examinees' use of English. To pass the writing test, examinees' essays for both task 1 and task 2 needed to receive band three or above in all of the four parts (see Appendix 2).

2. Pretest

A large-scale advanced level writing pretest was conducted in October 2001. Of the 234 examinees who took the advanced level listening and reading pretests, 102 took the writing pretest. Among them (see Table 2), group A consisted of thirty-four seniors majoring in English, the advanced level target examinees, from six public or private universities. Group B consisted of thirty-nine examinees who had passed the GEPT High-Intermediate level test. Group C consisted of twenty-nine examinees who had failed the GEPT High-Intermediate level test. All of the examinees were asked to fill out a questionnaire at the end of the pretest.

<table>
<thead>
<tr>
<th>Group</th>
<th>Background</th>
<th>Examinees (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Seniors majoring in English at university</td>
<td>34</td>
</tr>
<tr>
<td>B</td>
<td>Those who had passed the GEPT High-Intermediate level test</td>
<td>39</td>
</tr>
<tr>
<td>C</td>
<td>Those who had failed the GEPT High-Intermediate level test</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>102</td>
</tr>
</tbody>
</table>

The large-scale pretest was aimed at examining
- whether the standardized testing administration was appropriate;
- whether the revised prototype solved the previous problems;
- whether the revised rating scales were effective in discriminating examinees' abilities;
- whether the descriptors of the rating scales should be further modified;
- whether the rating standards were consistent;
- whether rating and score reporting procedures were appropriate;
- what cut score was appropriate.
II. Rating Process and Scoring Methods

A total of 102 valid writing samples were obtained in this pretest (tasks 1 and 2). Seventeen English experts, who were native speakers of either Mandarin or English, received rater training. To achieve rating consistency and stability, the GEPT team first familiarized the raters with the rating criteria and conducted trial rating. Then, twenty-four of the 102 writing samples were randomly selected in order to produce a marking scheme, which was verified by three of the GEPT committee members. Due to the small number of samples, only six of the seventeen raters were invited to do trial rating again with reference to the marking scheme.

In the formal rating process, each test paper was marked by two raters, and given scores had to be integers. Task 1 and task 2 were marked separately. All of the raters had to finish marking task 1 papers before marking task 2 papers to avoid the halo effect. The guidelines for calculating scores for each task are listed below:

1) The writing scores would not be converted or cumulated on the basis of a centesimal scale. Instead, examinees would get eight individual sub-scores, one half for task 1 and the other half for task 2, as shown in Table 3: task 1 "3, 3, 2, 2.5" and task 2 "3, 2.5, 2.5, 2."

2) When the discrepancy between the two total scores given in the double-rating process was less than 4 (< 4), and that discrepancy did not affect the "pass/fail" result, the final scores would be the average of the two sub-scores.

3) When the discrepancy between two total scores was more than 4 (> 4), or when the discrepancy between two sub-scores would affect the "pass/fail" result, a third rater would be asked to give the final scores. For example, as shown in Table 3, the discrepancy between the two total scores was more than 4 and that discrepancy would affect the "pass/fail" result. A third rater was asked to give the final sub-scores "3, 2.5, 2.5, 2" for the four parts of Task 2, respectively.

Table 3  Scoring of advanced level writing pretest

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance and adequacy</td>
<td>Coherence and organization</td>
</tr>
<tr>
<td>First rating</td>
<td>3</td>
</tr>
<tr>
<td>Second rating</td>
<td>3</td>
</tr>
<tr>
<td>Third rating</td>
<td>—</td>
</tr>
<tr>
<td>Final scores</td>
<td>3</td>
</tr>
</tbody>
</table>

1 Halo effect: a tendency for judges to assign similar scores across scales causing them to be more strongly correlated with each other. (Yorozuya and Oller, 1980)
III. Descriptive Analysis of the Pretest Results

1. Overall statistical results

The mean score, standard deviation, and standard error (SEM) of the 102 writing pretest examinees were 2.26, 0.539, and 0.54, respectively. The mean score and standard deviation of group A (seniors majoring in English at university) were 2.52 and 0.585, respectively. The mean score and standard deviation of group B (those who had passed the GEPT High-Intermediate level test) were 2.34 and 0.350, respectively. The mean score and standard deviation of group C (those who had failed the GEPT High-Intermediate level test) were 1.84 and 0.456, respectively. The pretest results showed a significant difference \((p < 0.01)\) in the scores between groups A and C and between groups B and C, indicating that this pretest form could discriminate well between strong examinees' writing ability (see Table 4).

<table>
<thead>
<tr>
<th>Group</th>
<th>Background</th>
<th>Examinees (N)</th>
<th>Mean score</th>
<th>Standard deviation</th>
<th>Highest score</th>
<th>Lowest score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Seniors majoring in English at university</td>
<td>34</td>
<td>2.52</td>
<td>0.585</td>
<td>3.69</td>
<td>1.75</td>
</tr>
<tr>
<td>B</td>
<td>Those who had passed the GEPT High-Intermediate level test</td>
<td>39</td>
<td>2.34</td>
<td>0.350</td>
<td>3.44</td>
<td>1.81</td>
</tr>
<tr>
<td>C</td>
<td>Those who had failed the GEPT High-Intermediate level test</td>
<td>29</td>
<td>1.84</td>
<td>0.456</td>
<td>2.38</td>
<td>1.19</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>102</td>
<td>2.26</td>
<td>0.539</td>
<td>3.69</td>
<td>1.19</td>
</tr>
</tbody>
</table>

Standard Error 0.54

*Groups A and B  \(p=0.07\)  Groups A and C  \(p < 0.01\)  Groups B and C  \(p <0.01\)*

2. Passing results

The cut score was originally planned as "3." By the standard, seven of the 102 writing pretest examinees achieved a mean score of 3 or above; six of the total examinees passed the test, which meant their sub-scores were all above 3. If the cut score was lowered to "2.75," which meant all of the mean sub-scores were equal to or above 2.75, thirteen of the total examinees would be deemed to pass the test (see Table 5 and Figure 1).

<table>
<thead>
<tr>
<th>Range of mean scores</th>
<th>Examinees (N)</th>
<th>%</th>
<th>Cumulative (N)</th>
<th>Cumulative %</th>
<th>Cumulative passing examinees (N)</th>
<th>Cumulative passing %</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.00 ~ 5.00</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>3.75 ~ 3.99</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>3.50 ~ 3.74</td>
<td>2</td>
<td>2%</td>
<td>2</td>
<td>2%</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>3.25 ~ 3.49</td>
<td>2</td>
<td>3%</td>
<td>4</td>
<td>4%</td>
<td>4</td>
<td>4%</td>
</tr>
</tbody>
</table>
Eighteen of the 102 pretest examinees passed the first-stage test by achieving scores of 80 in both listening and reading tests. With the eighteen examinees serving as the sample population, six of these examinees achieved band 3 in their writing test, and the passing rate was 33% (see Table 6), which was quite similar to the passing rates of the current GEPT Intermediate and High-Intermediate Level writing tests.

### Table 6  Number of examinees passing the GEPT Advanced listening, reading, and writing tests

<table>
<thead>
<tr>
<th>Examinees passing the listening and reading tests (N)</th>
<th>Examinees passing the writing test (N)</th>
<th>Passing rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>6</td>
<td>33%</td>
</tr>
</tbody>
</table>
When analyzed by groups, group A (main target examinees) had ten members passing the writing test, accounting for the largest percentage of the passing examinees, group B had three members doing so, and group C had no members doing so (see Table 7). The test results matched the expectations of the research team.

<table>
<thead>
<tr>
<th>Examinees by groups</th>
<th>Examinees (N)</th>
<th>Passing examinees (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Seniors majoring in English at university</td>
<td>34</td>
</tr>
<tr>
<td>B</td>
<td>Those who had passed the GEPT High-Intermediate level test</td>
<td>39</td>
</tr>
<tr>
<td>C</td>
<td>Those who had failed the GEPT High-Intermediate level test</td>
<td>29</td>
</tr>
</tbody>
</table>

3. Statistical results of writing test performance by groups

Ten of the thirty-nine group B examinees (who had passed the GEPT High-Intermediate level test) received scores in band 5 for their HI writing test, and their mean scores on the advanced writing test ranged between 1.9 and 3.4; thirty-three examinees who earned scores in band 4 for their HI writing test obtained mean scores of the advanced writing test ranging between 1.5 and 2.7. Nineteen of the twenty-nine group C examinees who earned scores in band 3 for their HI writing test obtained mean scores of the advanced writing test ranging between 1.2 and 2.4; five examinees who scored in band 2 or below for their HI writing test obtained mean scores of the advanced writing test ranging between 1.5 and 2.0 (see Table 8).

<table>
<thead>
<tr>
<th>Range of HI writing scores</th>
<th>Participating examinees (N)</th>
<th>Passing examinees (N)</th>
<th>Range of mean scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>90–100 (Band 5)</td>
<td>10</td>
<td>3</td>
<td>1.9–3.4</td>
</tr>
<tr>
<td>80–89 (Band 4)</td>
<td>33*</td>
<td>0</td>
<td>1.5–2.7</td>
</tr>
<tr>
<td>60–79 (Band 3)</td>
<td>19</td>
<td>0</td>
<td>1.2–2.4</td>
</tr>
<tr>
<td>40–59 (Band 2)</td>
<td>4</td>
<td>0</td>
<td>1.5–2.0</td>
</tr>
<tr>
<td>Below 40 (Band 1)</td>
<td>1</td>
<td>0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

* One examinee did not complete the test, and thus was not included in the statistical results.

Only three of the group B examinees who were in band 5 for their HI writing test passed the advanced writing test. The results indicated that the advanced writing test had an appropriate difficulty level and discriminated examinees of advanced level and high-intermediate level well, which met the original expectations of the research team.
4. Statistical results by item types

Each of the 102 examinees achieved similar results in both writing tasks. The overall mean score and standard deviation of task 1 were 2.28 and 0.63, respectively; the overall mean score and standard deviation of task 2 were 2.24 and 0.50, respectively. The $p$ value was 0.646, indicating both tasks 1 and 2 were of similar difficulty. The correlation coefficient between tasks 1 and 2 was 0.80, indicating that the abilities tested by the two tasks were relevant to each other but not identical (see Table 9).

<table>
<thead>
<tr>
<th></th>
<th>Task 1</th>
<th>Task 2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean score</td>
<td>2.28</td>
<td>2.24</td>
<td>2.26</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.63</td>
<td>0.50</td>
<td>0.54</td>
</tr>
<tr>
<td>Passing examinees (N)</td>
<td>24 (24%)</td>
<td>17 (17%)</td>
<td>13 (13%)</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p=0.646$

As shown in Figures 2 and 3, the score distribution of task 1 was platykurtic, suggesting a wider variation in the examinees' ability to summarize articles and express their opinions; the scores of examinees who passed task 1 were mostly clustered between 2.75 and 2.99. The score distribution of task 2 was leptokurtic compared with that of task 1, meaning that the examinees' ability to give a description, discuss possible causes, and make recommendations of non-verbal inputs was more similar to each other.

Figure 2  Score distribution of task 1
5. Statistical results of analytical scoring parts

According to analysis of Many-Facet Rasch Measurement (see Table 10), the overall mean scores of the four parts (relevance & adequacy, coherence & organization, lexical use, and grammatical use) in tasks 1 and 2 ranged between 2.5 and 1.9. The Separation value was 5.79 and the significance was .00. The Infit Mean Square values of the analytical scoring parts ranged between 0.8 and 1.1, all of which falling below parameter 1.2,\(^2\) indicating that the rating scales and the descriptors of the scales were appropriate and discriminating well between strong examinees’ writing ability. Also, as shown in Table 10, the examinees generally performed better on relevance & adequacy and coherence & organization (mean scores ranging from 2.3 to 1.9), which matched the expectations of the research team.

Table 10  Statistical results of analytical sub-scores for tasks 1 and 2

<table>
<thead>
<tr>
<th>Obsvd Score</th>
<th>Obsvd Count</th>
<th>Obsvd Average</th>
<th>Fair-M</th>
<th>Model Measure</th>
<th>Infit S.E.</th>
<th>Infit MnSq</th>
<th>Infit ZStd</th>
<th>Outfit MnSq</th>
<th>Outfit ZStd</th>
<th>N ITEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>590</td>
<td>233</td>
<td>2.5</td>
<td>2.52</td>
<td>.11</td>
<td>.12</td>
<td>1.1</td>
<td>1</td>
<td>1.1</td>
<td>1</td>
<td>5 5</td>
</tr>
<tr>
<td>591</td>
<td>235</td>
<td>2.5</td>
<td>2.50</td>
<td>.18</td>
<td>.12</td>
<td>1.1</td>
<td>0</td>
<td>1.1</td>
<td>0</td>
<td>2 2</td>
</tr>
<tr>
<td>573</td>
<td>235</td>
<td>2.4</td>
<td>2.42</td>
<td>.45</td>
<td>.12</td>
<td>0.9</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>1 1</td>
</tr>
<tr>
<td>548</td>
<td>233</td>
<td>2.4</td>
<td>2.33</td>
<td>.74</td>
<td>.12</td>
<td>1.1</td>
<td>0</td>
<td>1.1</td>
<td>0</td>
<td>6 6</td>
</tr>
<tr>
<td>530</td>
<td>235</td>
<td>2.3</td>
<td>2.22</td>
<td>1.08</td>
<td>.12</td>
<td>0.8</td>
<td>-1</td>
<td>0.8</td>
<td>-1</td>
<td>3 3</td>
</tr>
<tr>
<td>503</td>
<td>235</td>
<td>2.1</td>
<td>2.10</td>
<td>1.48</td>
<td>.12</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>4 4</td>
</tr>
<tr>
<td>488</td>
<td>233</td>
<td>2.1</td>
<td>2.06</td>
<td>1.63</td>
<td>.12</td>
<td>0.9</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>7 7</td>
</tr>
<tr>
<td>444</td>
<td>233</td>
<td>1.9</td>
<td>1.87</td>
<td>2.29</td>
<td>.12</td>
<td>1.1</td>
<td>0</td>
<td>1.1</td>
<td>0</td>
<td>8 8</td>
</tr>
</tbody>
</table>

533.4 234.0 2.3 2.25| .99 .12 | 1.0 -0.0 | 1.0 -0.0| Mean (Count: 8)
49.1 1.0 0.2 0.22| .72 .00 | 0.1 1.0 | 0.1 1.0| S.D.

RMSE (Model) .12 Adj S.D. .71 Separation 5.79 Reliability .97
Fixed (all same) chi-square: 273.8 d.f.: 7 significance: .00
Random (normal) chi-square: 7.0 d.f.: 6 significance: .32

\(^2\) The parameter was calculated by the formula: Mean + 2 S.D.=1.0 + (0.1×2)=1.2
6. Inter-rater reliability

According to Rasch analysis (see Table 11), the Separation value calculated by the scores given by six raters was 0.47, and the significance was 0.09, indicating that there was high consistency in using the rating scales among the six raters, and that there was no significant difference among the given scores. The inter-rater reliability was 0.82. In addition, all of the six Infit Mean Square, with values ranging between 0.6 and 1.2, fell below parameter 1.4, which showed there was great intra-rater reliability.

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obsvd</td>
<td>Obsvd</td>
</tr>
<tr>
<td>Score</td>
<td>Count</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1263</td>
<td>560</td>
</tr>
<tr>
<td>451</td>
<td>200</td>
</tr>
<tr>
<td>905</td>
<td>392</td>
</tr>
<tr>
<td>464</td>
<td>200</td>
</tr>
<tr>
<td>908</td>
<td>400</td>
</tr>
<tr>
<td>276</td>
<td>120</td>
</tr>
<tr>
<td>711.2</td>
<td>312.0</td>
</tr>
<tr>
<td>361.3</td>
<td>151.4</td>
</tr>
</tbody>
</table>

RMSE (Model) .12 Adj S.D. .06 Separation .47 Reliability .18
Fixed (all same) chi-square: 9.5 d.f.: 5 significance: .09
Random (normal) chi-square: 5.5 d.f.: 4 significance: .24

7. Results of the questionnaire survey

A total of ninety-seven valid questionnaires were received at the end of this pretest. The summarized results are presented in the following. For detailed contents and results of the questionnaire survey, please refer to Appendix 3.

Instructions for the writing test:

More than 80% of the questionnaire respondents considered the instructions for the writing test to be clear; up to 95% of the respondents considered the instructions for task 2 to be clear. Approximately 10% of the respondents felt confused about the instructions for task 1, where examinees had to finish reading two articles before responding to the questions, indicating that some of the examinees were not familiar with this relatively unconventional item type.

Time for responding to the questions:

85% of the respondents considered the amount of time provided for responding to the questions to be just right or slightly long. The remaining 15% of the respondents considered the time to be slightly short, which was a reasonable result as twenty-nine examinees who had failed the GEPT High-Intermediate level test participated in the questionnaire survey, accounting for 28% of the total respondents.

---

3 Table 11 was the report of FACETS statistical analysis, in which reliability referred to separation reliability (SR). SR was similar to KR20. The conversion between SR and IRR (inter-rater reliability) was calculated by the formula: IRR = 1 - SR = 1 - 0.18 = 0.82.

4 The parameter was calculated by the formula: Mean + 2 S.D. = 1.0 + (0.2×2) = 1.4
Perceived difficulty of item types:
More than 97% of the respondents considered the difficulty level of the verbal inputs in task 1 and that of the non-verbal inputs in task 2 to be just right, which suggested that reading ability was not among the variables that would affect examinees' writing performance. Approximately 70% of the respondents agreed that this test could adequately assess their writing ability; taken separately, more respondents answered affirmatively for task 1 (74%) than for task 2 (69%). The survey results matched the examinees' weaker performance on interpreting the non-verbal inputs.

Others:
In terms of the topic, text length or figures, layout, and space for writing drafts, more than 80% of the respondents considered this writing test to be satisfactory, and up to 95% of the respondents considered task 2 to be satisfactory. However, some of the respondents were not satisfied with the size of the font and the space for writing drafts. Most of them wished to have the test booklet enlarged from size B5 to A4.

IV. Conclusion and Suggestions

The results of the second-stage advanced level writing pretest were quite satisfactory. More specifically, the application and analysis results of the item types, rating scales, scoring methods, cut score, and testing procedures generally corresponded to the research team's expectations. Therefore, with minor modifications, the advanced level writing test will be ready to be formally administered. Suggestions for the present modifications and future research directions are as follows:

1. Modifying the band 5 descriptors for lexical use and grammatical use

None of this writing pretest examinees received scores in band 5. There might be several reasons behind this. First, group A examinees (major target examinees) were just becoming university seniors (at entrance-level) rather than English graduates (at exit-level) when the pretest took place in October. Second, group A examinees participated in the pretest on a voluntary basis, and perhaps none of them were top-notch university students. Third, the original band 5 descriptors for lexical use and grammatical use might have discouraged the raters from giving high scores. Therefore, modifications are suggested:

Descriptor 2 (original version):
- Vocabulary is **always** used appropriately.
- Structures are **always** used accurately and appropriately.

Modified version:
- Vocabulary is **always** used appropriately. *Errors are rare.*
- Structures are **always** used accurately and appropriately. *Errors are rare.*

The modifications were not only more consistent with use of language in real life but were also justified by the validity research of the Cambridge CPE (Certificate of Proficiency in English) test of the same level as the GEPT Advanced writing test.
2. Collecting writing samples in each band

Writing samples in various bands were collected following the end of the pretest (see Appendix 4). The research team will continue to collect special writing samples or those rated in different combination of bands in order to establish rating standards for the advanced writing test, which will serve as a basis for training and trial rating, producing marking schemes, and dealing with special cases.

3. Training raters

The procedures for training raters for this pretest, which are listed below, might be models for future tests:

- Familiarize raters with the scoring procedure by providing them with descriptors of advanced writing ability, test forms, explanations for rating scales, guidance notes on rating, and writing samples.
- Hold rater training sessions where rating principles can be explained and questions about rating answered.
- Conduct trial rating

4. Modifying the layout of the test booklet

The results of the questionnaire survey showed that approximately 15% of the respondents considered that there was not enough space for writing drafts. Therefore, the size of the test booklet should be enlarged from B5 to A4, which would leave more margins on the pages for writing task 1 drafts.

5. Providing more explanations and promoting the advanced writing task types

As indicated in the statistical analysis of writing tasks, scoring criteria, and questionnaire survey, the target examinees' unfamiliarity with the integrative (reading and writing integrated) testing format of task 1 might indirectly affect test performance. As for task 2, although more than 95% of the questionnaire respondents considered the design of task 2 to be satisfactory, the examinees performed worse on task 2 than on task 1. As a result, more explanations about these item types should be provided for examinees and educational institutes in the future to reduce examinees' anxiety.
V. Appendices

Appendix 1: The GEPT Advanced Level Writing Pretest form

GEPT Advanced Writing Test

**General Instructions:** In this test, you will have an opportunity to demonstrate how well you can extract main ideas from both verbal and non-verbal input, organize these ideas effectively in writing tasks and make clear your own viewpoint on these main ideas. There are two tasks in this test. Each task has a different time limit. You must complete both tasks or your test will not be marked. Detailed instructions will be given to you at the beginning of each task.

Read these instructions carefully and plan ahead so that you can complete all the requirements within the time limit. Insufficient development of any part of either task will result in a lower score for that task. Your performance will be scored according to the following criteria: relevance and adequacy, coherence and organization, lexical use, and grammatical use.

The entire writing test takes 105 minutes.
Task 1

You are going to take part in the GEPT Composition Contest and the winning prize is an NT$5,000 book voucher. The title of the composition is **The Advantages and Disadvantages of Advertising.** Information about this topic is provided in the two articles on the following pages.

- First, read each text to establish the main points the writer is making. You can use the space provided on your test paper to make notes.

- Then, **in your own words** as far as possible, write a composition that summarizes the **main ideas of both texts** concerning the pros and cons of advertising. If you use more than three consecutive words from the articles, use quotation marks (" "). **Plagiarism will result in failure.**

- In the final part of the composition, you should make clear your own viewpoint on these main ideas and come to a conclusion.

Your composition must be about **250 words.** You have **60 minutes** to complete Task 1.

**The Advantages and Disadvantages of Advertising**
The Disadvantages of Advertising

Anyone who lives in America is aware of advertising. Reading newspapers and magazines, we see full-page ads urging us to buy clothes, autos, cigarettes, and kitchen appliances. Television and radio programs include commercials; we all have heard the phrases "brought to you by" and "sponsored by" hundreds of times. If we drive, we see road signs or billboards proclaiming the qualities of products or the location of restaurants or motels. If we commute on a transit system, we cannot help but notice the prominent signs displayed on the buses and subways. And in our mail, along with the bills and the letters, come shiny flyers and circulars promoting products and announcing sales.

Advertisements in some form intrude into nearly every waking minute of our lives. We simply cannot get away from their pounding, incessant messages. Because ads permeate radio and television, we find ourselves singing their silly jingles and repeating their "cute" lines. Sellers admonish us to buy through a profusion of techniques: hard sell, soft sell, music, comedy, and appeals to all our emotions and fears.

Some ads are even potentially harmful. Perplexing or misleading sales pitches may lure unwary buyers into financial trouble. It is always best to remember: "caveat emptor"—let the buyer beware. Many commercials go far beyond the mere transmitting of information when they attempt to transform our values and attitudes. Cigarettes ads, for example, often imply that smoking is a manly or sexy habit. It is neither.

Because of these problems, many people have become extremely critical of commercials, especially those directed at children. As adults, we are often skeptical of what we read or hear in advertisements. Children, because they are not as mature or experienced as we are, cannot judge how reasonable or accurate ads are. If the man on TV says chocolate-covered, sugar-coated wheat toasts are healthful and nutritious, children may very well believe it. Many parents feel sellers take unfair advantage of children's inability to evaluate what they see or hear.

Unfortunately, despite their problems, commercials and ads are an established part of modern life. Providers of goods and services will always try to persuade us to purchase what they are selling. As consumers, we must learn not to believe everything we hear or read, so that we will not be fooled into buying things we don't need.

(Adapted from the article by Dennis Keen)

NOTES
The Advantages of Advertising

Advertising plays an extremely important role in our society. Perhaps most obviously, it keeps us informed about the latest products and services, thus enabling us to buy intelligently. Advertising also has a positive impact on our economy, by providing funding for the media and stimulating competition among goods and service providers. In addition, advertising can be used to promote public welfare, thus exerting a positive social impact on society.

We as consumers benefit greatly from advertising. By reading bank ads, for example, we might decide to transfer our money from our current bank to one offering better rates or more convenient hours. When traveling, we can save hundreds of dollars on transcontinental airfares by comparing the ads in the travel section of the newspaper.

Advertising can also be used to increase awareness in society about particular issues, and in so doing, it becomes a form of education. Anti-drug advertising such as "Just say NO", and drunk driving campaigns are just two examples of how society uses the advertising industry as a means to promote public welfare.

In addition to the social benefits, advertising also brings huge economic benefits to society. Without advertising, the media -- including newspaper, television, radio, etc. -- would be much less vigorous. Advertising provides revenue for commercial mediums which would otherwise need to be funded by the actual consumer of these mediums. For example, a newspaper would cost up to three times as much money (since advertising provides two-thirds of the revenue of the print media), or all television, bar government funded networks, would be pay-TV (since nearly all revenue for television is currently provided by advertising). The price a consumer may have to pay to receive very cheap, or even free, news and entertainment may include sitting through a 30-second commercial break while watching a television program, or flicking a couple of extra pages in a magazine, to get through the advertisements to the articles.

Although advertising might appear to raise the prices of goods and services, a closer look will show that it actually helps to keep prices low. Advertising stimulates economic activity, with vigorous competition between institutions and higher buying rates of products. This, in turn, leads to lower product costs for the consumer.

(Adapted from article found at: http://www.essayworld.com/members/essays/30/4020.shtml)

----------------------------------------------------------------------------------------------------------------------------------
NOTES

DO NOT TURN TO THE NEXT PAGE UNTIL YOU ARE TOLD TO DO SO.
Task 2

A local English newspaper has just printed some worrying statistics on the traffic accidents that occurred in the downtown area in June. The data are shown in Figures 1 and 2 below. As a citizen, you would like to help improve the situation.

Write to the Opinion section of this local English newspaper:

- Firstly, summarize what you think are the main findings from the reported data and discuss the possible causes.

- Secondly, make suggestions about what can be done to reduce the number of accidents in the downtown area.

Your report must be about 250 words. You have 45 minutes to complete Task 2.

Figure 1. Number of accidents in the downtown area in June

Figure 2. Percentage of accidents involving drivers in different age groups
## Appendix 2: Rating scales for the GEPT Advanced Level Writing Test

<table>
<thead>
<tr>
<th>Score</th>
<th>1</th>
<th>2</th>
<th>3 (Pass)</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Relevance and Adequacy | *text lacks relevance*  
  *parts of the task not addressed*  
  *nearly all main ideas from the input are missing*  
  *personal opinions are missing, not clearly stated or inappropriate to the task*  
  [pros & cons and conclusion, task 1; speculation on causes, and recommendations, task 2] | *almost entirely relevant*  
  *all parts of the task are addressed*  
  *may miss one or two main ideas from the input*  
  *personal opinions are appropriate to the task and clearly stated*  
  [pros & cons and conclusion, task 1; speculation on causes, and recommendations, task 2] | *entirely relevant*  
  *all parts of the task are addressed*  
  *all main ideas from the input are covered*  
  *personal opinions are appropriate to the task and effectively stated*  
  [pros & cons and conclusion, task 1; speculation on causes, and recommendations, task 2] | | |
| Coherence and Organization | *text lacks logical organization which may lead to confusion*  
  *inappropriate paragraphing*  
  *limited/inappropriate use of linking devices* | *text is logically organized in general*  
  *appropriate paragraphing*  
  *appropriate use of linking devices* | *text is logically organized throughout*  
  *appropriate paragraphing*  
  *wide and appropriate use of linking devices* | | |
| Lexical Use  
  *range*  
  *appropriacy* | *range of vocabulary is inadequate to complete the tasks*  
  *vocabulary is frequently used inaccurately*  
  *inappropriate/mixed register, showing that the examinee is unable to distinguish between registers*  
  *overt plagiarism [task 1]* | *adequate range of vocabulary is used to complete the tasks*  
  *there may be some inappropriate use of vocabulary*  
  *appropriate register with only occasional slips* | *wide range of vocabulary is used to effectively complete the tasks.*  
  *vocabulary is used appropriately. Errors are rare.*  
  *consistently appropriate register* | | |
| Grammatical Use  
  *range*  
  *accuracy* | *too limited a range of structures for task completion*  
  *sentences/structures are frequently used inaccurately and/or inaccurately* | *an adequate range of structures for task completion*  
  *there may be some inaccurate structures* | *effective use of a wide range of structures to complete the tasks*  
  *structures are used accurately and appropriately. Errors are rare.* | | |

* plagiarism: more than three consecutive words are copied from the input without appropriate quotation

### Non-ratable compositions

A composition is considered "non-ratable" when  
* the length is shorter than 80 words, or  
* the content is totally irrelevant, or  
* the content is apparently a copy or memorization from other known materials, or  
* the handwriting is completely unintelligible.

Non-ratable compositions will receive a final score of "0" and marks for the subcategories will not be reported.
Appendix 3: Results of the questionnaire survey

**Task 1**

26. The test instructions in Task 1 are:

- A. Clear [Bar chart showing 82%]
- B. Unclear [Bar chart showing 14%]
- C. Other [Bar chart showing 4%]

27. The two articles in Task 1 are:

- A. Difficult [Bar chart showing 1%]
- B. Just right [Bar chart showing 79%]
- C. Easy [Bar chart showing 18%]
- D. Other [Bar chart showing 2%]

28. The amount of time provided for responding to Task 1 is:

- A. Slightly long [Bar chart showing 6%]
- B. Just right [Bar chart showing 78%]
- C. Slightly short [Bar chart showing 16%]
29. In my opinion, **Task 1** can reflect my English **writing** ability:

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Agree</td>
<td>74%</td>
</tr>
<tr>
<td>B. Barely Agree</td>
<td>23%</td>
</tr>
<tr>
<td>C. Disagree</td>
<td>3%</td>
</tr>
</tbody>
</table>

30. To me, the topic, text length, layout, and space for writing drafts in **Task 1** are:

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Just right</td>
<td>80%</td>
</tr>
<tr>
<td>B. Other</td>
<td>19%</td>
</tr>
</tbody>
</table>

**Task 2**

31. The test instructions in **Task 2** are:

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Clear</td>
<td>95%</td>
</tr>
<tr>
<td>B. Unclear</td>
<td>5%</td>
</tr>
</tbody>
</table>

32. To me, it is difficult to interpret the figures in **Task 2**:

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. No</td>
<td>97%</td>
</tr>
<tr>
<td>B. Yes</td>
<td>2%</td>
</tr>
<tr>
<td>C. Other</td>
<td>1%</td>
</tr>
</tbody>
</table>
33. The amount of time provided for responding to Task 2 is:

- Slightly long: 11%
- Just right: 74%
- Slightly short: 16%

34. In my opinion, Task 2 can reflect my English writing ability:

- Agree: 69%
- Barely agree: 28%
- Disagree: 2%

35. To me, the topic, figures, layout, and space for writing drafts in Task 2 are:

- Just right: 95%
- Other: 3%
- No answer: 2%
Appendix 4: Advanced level writing samples

Task 1: writing sample 1

Whenever we read newspapers and magazines, we can see different kinds of ads including clothes, autos, cigarettes, and so on. Some unwary buyers may be lured into financial trouble because of some potentially harmful ads. Moreover, the same ad has different influence on adults and children. Adults can tell what is reasonable and accurate ads from the ads, but children cannot do it.

However, advertising can also be beneficial to us. It has several advantages for us and our society. First of all, it can bring huge economic benefits to society. Second, taking anti-drug advertising for example, it can be a form of education. Last but not least, advertising has a positive impact on our economy.

To sum up, whatever advantages or disadvantages of advertising are, we should "learn not to believe everything we hear or read, so that we will not be fooled into buying things we don't need."

READER'S COMMENT

Non-ratable composition  R&A  O&C  LU  GU

The composition is far from fulfilling the 250-word length requirement. In addition, almost half of its content is copied from the input without quotation (e.g. ...read newspapers and magazines..., ...unwary buyers...into financial trouble, ...can bring huge economic benefits to society, ...has a positive impact on our economy, etc.). This makes the rest of the composition too short for the reader to properly evaluate the writer's knowledge of grammar and vocabulary.
Advertising has a great power in our daily life. It is obviously that whether the effect of ads is useful or not, it really into our couture. People who live in modern life are always forget that advertising is base on money. That is also the biggest disadvantages of ads. For the purpose of making money, advertisements usually tell a lie. When people believe it, they will get into trouble.

On the other hand, the advertising's media everywhere in our life, so if the ads can bring benefits to us, it has the education function. People can know many knowledge from the media. So government usually use ads to tell the new law or others. We can easily get these informations from Advertising.

In my opinion, I think that ads is very important. Try to image that if there are no ads in our daily life, we can't know many new production's information, and can't get the benefits from ads. However, life today is not easy, we must learn to choose the information from advertising. Everyone should have these kinds of ability. Than We can live well with advertising.

READER'S COMMENT

<table>
<thead>
<tr>
<th>Non-Pass 1</th>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L</th>
<th>G-U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The writer attempts but fails to summarize the main points from the input, nor does he/she succeed in expressing his/her personal opinions clearly. The organization is poor, often leading to confusion. The lexical range is limited, and there are serious errors in basic sentence structure.
Task 1: writing sample 3

In the free-trade of modern time, advertising is a must in every walks of life. And yet, it could become a nightmare if we buy whatever we see in ads just because it's said to be cheap, healthy, and fashion.

Advertising could be convenient and "a form of education", when it is served as a government's tool, telling people the right way to deal with drugs and drinkings, for example. On the other hand, unawared harmful or sexual commercials are held wide-open, teaching and guiding our children and generations to come the wrong way!

Therefore, we should keep our eyes open, while using the benefit and bargain of advertising try not to forget that "don't believe everything we hear or read," or we might "buy things we don't really need."

In my opinion, I think advertisements and commercials are great challenge of salesmen's intelligence. Once in a while, they have to produce new products to attract buyers attention. And they have to come up with neat words and fascinating ideas and phrases to persuade us into buying their stuff. "I'm not as interested in what's the present inside as in the wrapping paper." It's something to mention about. But needless to say, I hated when I open the mailbox and see it cramed with junk mails and only junk mails. It's a waste of environmental resource, and we don't keep ads in mass-commi such as the Web, TVs, and radios; and publish less newspaper and magazines? In addition, the authority also should make sure that the flood of ads wouldn't effect our children on the bad part.

READER'S COMMENT

Non-Pass 2

<table>
<thead>
<tr>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Both the summary and the personal opinion parts of the task are sufficiently addressed; however, some points are unclearly stated, especially in the part for personal opinions. Also, the last opinion is poorly developed and the composition ends abruptly. Overall, the paragraphing is appropriate, but the development of the ideas is awkward or unclear sometimes. The range of vocabulary and structures is not quite sufficient to complete the task and there are quite a few errors in their usage. Several instances of inappropriate register also appear in the composition (e.g. And yet,... and ...buying their stuff. etc.).
Nowadays, advertisements and commercials are an important part of our lives. We see them everywhere, on television, in newspapers or magazines, on public transportations and even on the streets. Advertising has great influence on us, and there are both advantages and disadvantages.

Advertising is simply trying to persuade consumers into buying something. So in order to sell their products, some advertisements may be dishonest and consumers often will be fooled. Ads may also have bad influence on children, since children are most likely to believe anything. Being a smart consumer, we should think twice before buying a product. After all, money is hard to earn.

On the other hand, advertising also has some advantages. We can get the latest informations of new products or services, so we could spend money on what we really need. Consumers can also compare different ads on a product to choose the best brand, and it may save a lot of money.

We often see lots of commercials on TV. Personally, I like to watch TV commercials very much. I think commercials are full of great and interesting ideas. Some ideas may be funny and humorous, and some commercials might even have jingles or slangs that are easy to remember. All of this just to attract more consumers. I think a lot of their ideas are very cute, so commercials might as well be an art.

We have to admit, advertising plays an important role in our daily lives. Although there are disadvantages and even may have bad influences, but as long as we think smart and don’t take everything in, there will be more advantages than we think.

**READER'S COMMENT**

<table>
<thead>
<tr>
<th>Pass 3</th>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Both the summary and the opinion parts of the task are sufficiently and clearly addressed. The composition is, in general, logically organized although there is an unexpected personal comment on a disadvantage in the middle of the summary. The paragraphing and use of linking devices is mostly appropriate. The range of vocabulary and structures is adequate, with only some minor errors in their usage and several instances of run-on sentences.
Advertising is crucial to our everyday life. It provides valuable information about the latest products, thus saving consumers time and money to get what they want. It also contributes a main part of the media revenue, enabling program producers to broadcast good and relatively inexpensive service to their audience. Needless to say, some advertising that targets at public good bring an enormously positive impact on the society.

However, some people view advertising as nothing but evil. They see advertising with its silly marketing phrases as an annoying intruder to our daily tranquility. In their eyes, advertising lies and exaggerates about products, exploiting numerous consumers, especially gullible and innocent children. Some advertisements, like those about cigarettes, even promote ideas that are harmful to public health.

In my own opinion, advertising has far more advantages than disadvantages. And we simply cannot live without it. Like a bridge, it helps manufacturers to get their product information across to the consumers. Without this channel, consumers will find out the best bargain in a later time and simply pass by the great opportunity because of inaccessibility to information. Therefore, advertising is an indispensable facilitator of our life.

As for the disadvantages of advertising mentioned by its opponents, I believe all of them can be corrected and even can be turned into a positive force. For example, in face of unreliable or exaggerative advertising, it is a great opportunity to educate our children not to believe what they see or hear easily. The world is full of lies. If the younger ones can learn to see through the pretended benevolence of many things through advertising then those false advertisements are great teachers in a way.

In terms of the annoying nature of advertising, I simply disagree. Those advertising cannot deny their function in spicing our daily conversation and providing us many laughs in our over stressful life.
Task 2: writing sample 1

From the data, we can find that traffic accidents take place mostly during rush hours and midnight. We must focus on the midnight in particular. Speeding and drunk driving usually happen in the period of time. And that's

**READER'S COMMENT**

<table>
<thead>
<tr>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-ratable composition

The composition is incomplete. It does not meet the minimum length requirement of 80 words. The reader is unable to evaluate the writer's language ability.

Task 2: writing sample 2

In the figure 1, we can know most accidents happened during 22:00~2:00; on the other hand, this figures also shows that it is safer time during 2:00~7:00. Morevver, we can know that ages of accident involving drivers are most under 21.

In my opinion, time during 22:00~2:00 is our sleep time. Therefore, we should be very careful during this time if we are driving. If we are asleep, we shouldn't keep driving. Moreover, some drivers thinks there are fewer cars on the road during 22:00~2:00, so they won't obey traffic rules and drive very fast. I think it is the main reason that resulted in accident. Then, most youngers under 21 age do not get the driver license or do not know traffic rules very clearly, so accident happened easily to them. If they put more emphasis on their lives, they would drive carefully.

**READER'S COMMENT**

<table>
<thead>
<tr>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-Pass 1

The length of the composition is far below the 250-word requirement. The summary of the information provided in the charts and the speculation of the causes are inadequate, and there is hardly any suggestion for reducing the number of car accidents. The organization is awkward and unclear. The writer does not have commend of sufficient vocabulary or structures to complete the task.
From the first figure, I find that during 22 o'clock to 2 o'clock of the day, the number of the traffic accidents is the highest. And from the second one, I find that the age group of under 21 causes the most traffic accidents. As the first figure is concerned, I make three possibilities that may cause the car accidents. First, during 7 o'clock to 10 o'clock of the day, some people may be late for work. So they drive more fastly than usual. Then it causes the car accident happen. Second, during 16 o'clock to 19 o'clock, people may be impatient to get into the traffic jams so they don't notice the traffic rules and the car accident happens. Third, during 22 o'clock to 2 o'clock, people maybe feel so tired and can not concentrate to drive. Because of this reason, the car accidents happen. As far as the second figure is concerned, people who are under 21 cause more car accidents. I feel that people who are under 21 like to drive high speed and often and neglect the traffic rules.

In my opinion, I think during the working hours, we can inform people of taking buses or MRT to decrease the number of cars. For the figure 2, I think the government should increase the age limitation to twenty-one-year old. Maybe in these ways, the number of the car accidents can be decreased.
We can easily see two apparent findings from figure one and two. First, we find three time periods when most accidents happen. Two of them are the rush hours during the day, when people are going to work or school, and when people are getting home. The other time is pretty late at night. In figure two, we find that nearly half of the drivers are under twenty-one. In the words, half of the traffic accidents are caused by these young drivers.

For each findings, we can find some possible causes. First, most accidents during the rush hours may be a result of the heavy traffic. Everyone is busy and in a hurry. And the drivers may become impatient in the traffic. But why do some many accident occur in the late night when the traffic is not as heavy? Perhaps, some drivers may become careless just because of the good traffic. Or, darkness may be another reason. Still, the most likely reason is drunk driving. Secondly, there are some reasons accounting for young drivers' accidents. Maybe, they are usually more impatient. Or, they have less experiences so that they are not used to the traffic so well. Or, some young drivers are too confident in their driving skills which are not as skillful as they think. These are some possible causes for this traffic accidents.

Maybe we can learn some lessons from the data and these possible causes. First, we should try to reduce the number of cars in the downtown area during rush hours. To encourage more people to take mass transportation, the mass transportation system and the service should be improved. Severe punishments should be strictly enforced. Thirdly, the standards of giving drivers licenses should be more strict. We have to make sure every drivers on the roads are well-trained. If we can do all these mentioned, a large number of accidents can be increased.

READER'S COMMENT

<table>
<thead>
<tr>
<th>Pass 3</th>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

All parts of the task are sufficiently and clearly addressed. In general, the paper is logically organized, with appropriate paragraphing and use of linking devices. The range of vocabulary and structures is adequate, and their usage is mostly correct. However, there are still some errors (e.g. But why do some many accident occur...).
These days two figures concerning the traffic accidents that occurred in the city center in June were released. The first figure is about the number of accidents in different times of the day. We find that the most dangerous traffic hours were between 10 PM and 2AM, with 40 cases reported, almost 1.3 causes a day on average. Obviously, traffic accidents take place most frequently when tired people are done with their partying late at night.

The second figure shows the relation between traffic accident occurrence rate and age groups. Interestingly, almost half (49%) of the traffic accidents were caused by the youth group under 21 years of age. Most drivers of this age are inexperienced and maybe some of them drive behind the wheel without a license. On the contrary, those over 50 years of age are more cautious and experienced, thus resulting in the least traffic accidents (5%).

From the statistics, we could infer that the high occurrence of traffic accidents late at night in Figure 1 was probably caused by the riskiest drivers – age group under 21 – in Figure 2. Night hours are the time when the young are most active. Maybe after a bottle or two in a pub, they simply are too high to follow the traffic rules strictly.

Therefore, our police should make more rounds of patrol near pubs or clubs where most youngsters congregate at night. They should make sure that nobody out of the pub will be drunken – driving home. The drunk partiers should always go home by their friends' cars or taxis. The presence of the police will remind these drivers to go home in a safer way.

Meanwhile, the police should also promote the idea of driving with caution and with a license. Many of the riskiest drivers (under age 21) probably don't have a driver's license. Our police staff should enforce strictly to get those unqualified drivers out of the driver's seat. By so doing, our city will be a safer one.

**READER'S COMMENT**

<table>
<thead>
<tr>
<th>Pass 4</th>
<th>R&amp;A</th>
<th>O&amp;C</th>
<th>L U</th>
<th>G U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The writer addresses all parts of the task, effectively summarizing the main findings from Figure 1 and Figure 2. He/she also discusses some possible reasons behind the findings, and suggests some ways to reduce the number of accidents. The paper is well-organized, with appropriate paragraphing. The use of vocabulary and structures is more than adequate and mostly appropriate. However, there are still some minor errors (e.g. "These days two figures...were released." and "drunken-driving home") which do not affect comprehension.
General English Proficiency Test
Advanced Level Speaking Test

Contents

I.  Pretest Development Process................................................................. S-2

II. Pretest Process..................................................................................... S-4

III. Descriptive Analysis of the Pretest Results......................................... S-5

IV.  Conclusion and Suggestions............................................................. S-8

V.  Appendices.......................................................................................... S-10
    Appendix 1: The speaking test format................................................. S-10
    Appendix 2: Holistic rating scales......................................................... S-14
    Appendix 3: Analytical rating scales.................................................... S-15
    Appendix 4: The survey questionnaire.................................................. S-16
    Appendix 5: Transcripts of successful examinees' speaking samples........ S-18
I. Pretest Development Process

Examinees who pass the GEPT advanced level speaking test can speak English fluently with only occasional errors and can express their opinions in academic or professional fields. The Advanced Speaking Test consisted of three parts: Warm-up Interview, Information Exchange, and Presentation. This structure was determined on the basis of the results of a three-stage pretest in which fifty-six examinees participated, and a revision of item types conducted from June to October 2001. The analysis of the third stage pretest is presented in the following:

1. Item Types

Table 1 shows the task type, interaction pattern, test focus, and task length of each part in the speaking test.

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Interaction Pattern</th>
<th>Test Focus</th>
<th>Length in Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1 Warm-up Interview</td>
<td>Interlocutor interviews candidates</td>
<td>Giving personal information and quick opinions</td>
<td>About 5 minutes</td>
</tr>
<tr>
<td>Part 2 Information Exchange</td>
<td>Interlocutor presents an information exchange task to the candidates for discussion</td>
<td>Exchanging information and opinions</td>
<td>About 7 minutes</td>
</tr>
<tr>
<td>Part 3 Presentation</td>
<td>Interlocutor presents a topic to each candidate</td>
<td>Expressing and elaborating opinions in extended turns</td>
<td>About 10 minutes</td>
</tr>
</tbody>
</table>

1) Test format

The duration of the test was approximately twenty-two minutes. Two or three examinees participated in one test at a time. The test was tape-recorded and videotaped throughout, and there was one interlocutor and one assessor during the test session. The interlocutor presided over the test by asking examinees questions (see Appendix 1 for the detailed test rubrics for interlocutors) and marked the examinees' performance based on the holistic rating scale. The assessor's duty was to mark the performance based on the analytical rating scale.

2) Test type

- Part I was a warm-up interview between the examinees and the interlocutor, which lasted for approximately five minutes and aimed to assess examinees' speaking ability to introduce themselves and answer questions.
- Part II was a seven-minute session, during which examinees would exchange information, discuss a topic with each other, and answer questions posed by the interlocutor. Part II aimed at assessing examinees' speaking ability to interact and discuss with others.
- Part III was a ten-minute session, during which examinees would take turns presenting a talk after thinking about a given question for two minutes. When one examinee finished giving a talk, the other examinee would need to summarize the presentation. Part III aimed at assessing examinees' ability to give and summarize presentations within a limited time.
3) **Item writing principles**

On the basis of the can-do descriptors, the research team created items that mainly aimed to elicit oral responses from examinees. With reference to the study of O'Sullivan, Weir & Saville (2001), Table 2 shows the language functions required by this test.

![Table 2](image)

4) **Rating standards and method**

A. **Rating standards**

The merits and demerits of analytical and holistic rating have been discussed in previous studies (Hughes, 1989; Bachman & Palmer, 1996). This speaking test adopted both holistic and analytical ratings. Two five-band (1, 2, 3, 4, 5) rating scales were used separately, and the cut scores were band 3. That the two rating methods were used at the same time could help verify final scores (Hughes, 1989: 110) and enhance inter-rater
reliability. For the holistic rating scale, please refer to Appendix 2. The analytical rating scale consisted of six criteria, and the definition of each standard is listed as follows:

- Pronunciation: accuracy of pronunciation and intonation
- Relevance & Adequacy: sufficiency and appropriacy of speech content
- Lexical Use: appropriacy and range of lexical use
- Grammatical Use: accuracy and range of grammatical use
- Fluency: fluency in oral use of language
- Coherence: coherence in long talks

For the analytical rating scale, please refer to Appendix 3.

B. Rating method
To achieve rating consistency and stability, interlocutors and assessors need to attend standardization sessions, which would familiarize them with the rating standards and descriptors of the rating bands. Interlocutors would use the holistic rating scale, while assessors used the analytical rating scale. In principle, to pass the advanced level speaking test, examinees had to obtain band 3 or above on the holistic rating scale, and each of their analytical scores had to be in band 3 or above on the analytical rating scale. If there was discrepancy between the scores on the two rating scales, interlocutors and assessors had to reach a consensus through discussion. If not, a third rater would give the final scores by reviewing a video recording of the test session.

II. Pretest Process
This pretest was aimed at determining:
- the clarity of rubrics
- the sufficiency of time allocated
- the usefulness of rating scales

1. Examinees
Fourteen people, seven males and seven females, took the pretest held in October 2001. The examinees were all target candidates of the advanced level test, consisting of the general public, people who had passed either the second stage, or the first stage only of the GEPT High-Intermediate Level test, and university seniors majoring in English. The sample size was quite small since the test format (interviews) as well as qualitative and quantitative analysis of test and rating performance consumed a lot of manpower and time.

2. Questionnaire survey
Each examinee was asked to fill out a questionnaire at the end of the speaking test. The questionnaire consisted of questions concerning examinees' background information, test format, the interlocutor's performance, the amount of time provided for responding to each part of the test, examinees' performance, perceived item difficulty, and face validity of the test.
3. Rating process and rating alignment

1) Stage 1

An interlocutor and an assessor would separately rate test performance at the end of the test. The two raters needed to discuss discrepancy in their scores if the discrepancy affected the "pass/fail" result (for example, if an examinee's performance was rated band 3 by the interlocutor and band 2 by the assessor). The cases where no consensus was reached between the two raters would be left to the rating alignment meeting.

2) Stage 2

A rating alignment meeting was held, in which participants were asked to give holistic and analytical scores, respectively, to some examinees' performances. Participants would also discuss further controversy over score discrepancies arising in stage 1 and give final scores under a consensus. Standard speaking samples in each band would be determined in the meeting and used in the future to train raters.

III. Descriptive Analysis of the Pretest Results

1. Score distribution

Six of the fourteen pretest examinees passed the speaking test. The mean score derived from holistic rating was 2.71. Table 3 shows the test performance of different groups.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Examinees (N)</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>General public</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>HI examinees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Those failing the second stage test</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Those passing the second stage test</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>English majors</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Total examinees (N)</td>
<td>14</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

As shown in Table 3, the two examinees who had failed the second stage of the GEPT High-Intermediate test did not pass the advanced level speaking pretest, while only one of the four examinees who had passed the second stage of the GEPT High-Intermediate test passed this pretest. Therefore, it could be inferred from the above results that the advanced level speaking test could effectively discriminate strong examinees' English ability. In addition, of all the advanced level speaking pretest examinees, ten took the first stage of the advanced level test (listening and reading) as well, and three of them passed it. Two of these three examinees passed the advanced level speaking pretest, while the other seven who had not passed the first stage of the advanced level test failed the speaking pretest. Therefore, it was inferred that the first stage of the advanced level test served the screening purpose well.
Moreover, the Many-Facet Rasch Measurement results showed that this speaking test could effectively discriminate examinees' ability (see Table 4, where examinees/candidates were arranged in order of ascending test performance). As shown in the column of RATERS, the rating of the two raters, marked as H (holistic) and A (analytical) respectively, were of similar level of severity. As shown in the "criteria" column, there was not much difference between the six rating criteria; however, the examinees performed best on pronunciation (Pron.), then fluency (Flu.), relevance & adequacy (Re. & Ad.), coherence (Co.), grammar, and lexis, which matched the research team's expectations.

Table 4

<table>
<thead>
<tr>
<th>Measr candidate</th>
<th>RATERS</th>
<th>criteria</th>
<th>IS.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measr candidate</td>
<td>RATERS</td>
<td>criteria</td>
<td>IS.1</td>
</tr>
<tr>
<td>high</td>
<td>severe</td>
<td>severe</td>
<td></td>
</tr>
<tr>
<td>13 H A</td>
<td>+(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 E</td>
<td></td>
<td>Grammar</td>
<td></td>
</tr>
<tr>
<td>3 C</td>
<td></td>
<td>Lexis</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Co.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>Flu.</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>H</td>
<td>Pron.</td>
<td></td>
</tr>
<tr>
<td>-1 L</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4 E S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-9 T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td></td>
<td></td>
<td>+(1)</td>
</tr>
<tr>
<td>high</td>
<td>(lenient)</td>
<td>easy</td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>lenient</td>
<td>easy</td>
<td></td>
</tr>
</tbody>
</table>

Co. = Coherence, Flu. = Fluency, Re. & Ad. = Relevance & Adequacy
Pron. = Pronunciation

2. Inter-rater reliability

After the two-stage rating alignment process, the inter-rater reliability between the two raters was 0.94, indicating a great rating consistency.

3. Examinees' responses

The results of the examinees' questionnaire survey are summarized as below (for the complete version of the questionnaire, please refer to Appendix 4):
Participation in other large-scale tests: Three of the fourteen pretest examinees had taken the TOEFL test before and another four had taken the TOEIC test before. One of the TOEFL examinees and one of the TOEIC examinees failed this speaking test (see Table 5).

<table>
<thead>
<tr>
<th>Code of examinees</th>
<th>TOEFL</th>
<th>TOEIC</th>
<th>GEPT advanced speaking pretest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
<td>Fail</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>980</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>915</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>617</td>
<td>905</td>
<td>✓</td>
</tr>
<tr>
<td>D</td>
<td>595</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>E</td>
<td>690</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>F</td>
<td>260 (CBT)*</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

*TOEFL CBT 250-260 ≈ TOEFL PBT 600-617

Test format: All of the fourteen pretest examinees considered the face-to-face interview to be friendly, natural, interesting, and professional. None of them considered their performance to be affected by the audio recorder or the video camera. Most of them considered the arrangements of the desks, chairs, and lighting to be satisfactory.

The interlocutor: Two of the examinees considered the speech rate of the interlocutor to be slightly slow, and the remaining examinees considered the speech rate to be just right. All of the examinees considered that the interlocutor was friendly and gave clear instructions.

Time for responding to the questions: Ten of the examinees considered the amount of time provided for responding to questions of each part to be just right. Another three examinees considered that the time provided for preparing Part III tasks should be thirty seconds to one minute longer. The one remaining examinee considered there should be more time provided for responding to Part I and Part II questions and preparing Part III tasks.

Perceived item difficulty: Most of the examinees considered Part III to be the most difficult as they were "unfamiliar with" the questions, not capable of "summarizing others' talks" or "nervous." Three of the examinees considered Part I to be the most difficult as they could not predict what questions would be asked. Meanwhile, most of the examinees considered Part I to be the easiest as Part I questions involved "everyday life" and "personal experience", were "more relaxing," or "could be prepared in advance." Some considered Part II to be the easiest as the questions allowed examinees to "freely express their opinions" or just required them to ask each other "numbers." However, Part III was considered to be the easiest by some other students as they "had practiced oral summarization or presentation in English listening classes before."

Face validity of the test: All of the examinees agreed that this test could adequately assess their speaking ability.
IV. Conclusion and Suggestions

The results of the pretest were generally satisfactory, but minor modifications to the item types and the scales were suggested below. The suggestions were put forward and approved in the twenty-seventh GEPT research committee meeting and would be adopted in the first formal advanced level test held in 2002.

1. Task types

1) Part I

Originally, examinees were allowed to introduce themselves freely for one minute. It was suggested that Part I be a guided self-introduction, where the interlocutor would give examinees topics of self-introduction, in order to prevent examinees from memorizing responses before the test. For example, the interlocutor might ask examinee A "Please tell us about your hobbies" and ask examinee B "Would you please tell us about your job?" instead of just giving the question "Would you please introduce yourself?"

2) Part II

- The "final decision" in Part II discussion should be deleted as the results of the pretest showed that the examinees simply repeated what was said in the discussion, which was unnatural.
- The interlocutor should pose open-ended questions to both examinees and let them discuss with each other rather than let them answer different questions separately. Originally, Part II/A allowed more turn-takings but shorter turns, while Part II/B allowed longer turns but fewer turn-takings. Therefore, the modification should improve the imbalance and enable strong examinees to demonstrate their speaking ability.

Table 7 Original Version and Revised Version of Part II

<table>
<thead>
<tr>
<th>Part</th>
<th>Version</th>
<th>Original Version</th>
<th>Revised Version</th>
</tr>
</thead>
</table>
| Part II | A | Information exchange  
Discussion about the advantages and disadvantages | Information exchange  
Discussion about the advantages and disadvantages |
| | B | Follow-up Questions from the Interlocutor:  
- B's final decision  
- A's final decision  
- Interlocutor asks A one topic-related question  
- Interlocutor asks B one topic-related question | Follow-up Questions from the Interlocutor:  
- Interlocutor asks both A & B two to three topic-related & open-ended questions for discussion |

- Non-verbal inputs might be added to the original "Information Gap" inputs to guide examinees' discussion and add variety to the contents of this item type.

3) Part III

The Part III questions should be consistent in the abstract/concrete level to avoid affecting examinees' performance. For example, the Part III questions in this pretest were "Is it ethical to use animals in laboratory tests?" and "Should more nuclear power plants be built in Taiwan?" It turned out that those examinees who were given the first question did not
perform as well as those given the second question. The possible reasons might be that the first question involved ethical values judgment and was more abstract than the second question, and that examinees' familiarity /unfamiliarity with topics might affect their test performance.

2. Rating scales

The observations and results of the pretest showed that even the strongest examinees still made errors in lexical or grammatical use and had brief pauses while speaking. Similar results were found in the validation study of the Cambridge CPE (Certificate of Proficiency in English) test as well. Therefore, the suggested modifications to the band 5 descriptors were presented in the following:

1) Holistic rating scales
   - The sentence "The candidate speaks with ease" should be changed to "The candidate speaks fluently with minimal hesitations."

2) Analytical rating scales
   - Lexical Use: The sentence "Lexis is always used appropriately" should be changed to "Lexis is used appropriately. Errors are rare."
   - Grammatical Use: The sentence "Errors are rare" should be added to the original descriptor.
   - Coherence, a sub-analytical part originally applied to Part I and Part III, could be applied to the modified Part II item type as well.
V. Appendices

Appendix 1: The speaking test format

<table>
<thead>
<tr>
<th>Interlocutor</th>
<th>Advanced Two-candidate Format</th>
<th>Part I</th>
<th>(approx. 5 minutes)</th>
</tr>
</thead>
</table>

Interlocutor (the interlocutor and the candidates are all standing up) Good morning (afternoon/evening). How are you both doing this morning (afternoon/evening)?

My name is ________ and this is my colleague, ________. (point to the assessor and pause for 3 seconds) He/She is going to be listening to us.

Your names are …? (point to both candidates)

【Candidate A】 and 【Candidate B】. (shake hands with each candidate and gesture the candidates to sit down, then the interlocutor and the assessor sit down as well)

First of all we'd like to know something about you, so I'm going to ask each of you to introduce yourself for one minute. I will stop you when your time is up. Now, 【Candidate A】 (point to Candidate A), please tell us about your hobbies.

Candidate A ☑ (After 1 minute)

Interlocutor Thank you. 【Candidate B】 (point to Candidate B), would you please tell us about your job?

Candidate B ☑ (After 1 minute)

Interlocutor Thank you! Now, I'm going to ask each of you a few questions about yourselves.

【Candidate B】 , …. (Select an appropriate topic area and ask the first question in that topic)

Candidate B ☑ (After 20–30 seconds)

Interlocutor Thank you. 【Candidate A】 , …. (Ask Candidate A the same question)

Candidate A ☑ (After 20–30 seconds)

Interlocutor Thank you. Now, 【Candidate A】 , …. (Ask the next question on the list)

Candidate A ☑ (After 20–30 seconds)

Interlocutor Thank you. 【Candidate B】 , …. (Ask Candidate B the same question)

Candidate B ☑ (After 20–30 seconds)

Interlocutor Thank you. Now, 【Candidate B】 , …. (Ask the following question on the list)

Candidate B ☑ (After 20–30 seconds)

Interlocutor Thank you. 【Candidate A】 , …. (Ask Candidate A the same question)

Candidate B ☑ (After 20–30 seconds)

Interlocutor Thank you.

<End of Part I>
Interlocutor | Now, you two are going to work together to complete a task. 
(Pass out the visual prompt cards to the candidates. Candidates have different 
prompt cards on the same types of cell phones.)
Please do not show your card to your partner.

Here's the situation. Each of you is going to buy a cell phone for yourself.
【Candidate A】，you have information on the three cell phones and 
【Candidate B】，you have different information on the same three cell 
phones. You will first need to exchange information to fill the chart. You may 
write on your card.

You will have two minutes to exchange information. I will stop you when 
your time is up. Now, please begin and 【Candidate A】，please speak first.

<Visual Prompt> | (See test question paper.)

Candidates A & B | ☺ (After 2 minutes)

Interlocutor | Thank you.
Now, during the next two minutes, you and your partner are going to discuss 
both the advantages and disadvantages of each phone. Remember, this is a 
discussion. Focus your talk on the advantages and disadvantages of each cell 
phone. I will stop you when your time is up. Now, please begin your 
discussion and 【Candidate B】，please speak first.

Candidates A & B | ☺ (After 2 minutes)

Interlocutor | Thank you. Now, 【Candidate A & B】，I would like you two to talk about the 
following question. (Select one question from the test paper & read it to the 
candidates.) You have one and a half minutes to discuss. I will stop you when 
your time is up.

Candidate A | ☺ (After 1 minute)

Interlocutor | Thank you. Now, 【Candidate A & B】，(Select another question from the test 
paper & read it to the candidates). You have one and a half minutes to discuss. I 
will stop you when your time is up.

Candidate B | ☺ (After 1 minute)

Interlocutor | Thank you.

<End of Part II>
In a few seconds, I will give each of you a card with a question on it. Each of you will have two minutes to think about what you are going to say. Then, you will have two minutes to present your talk to your partner. Listen carefully while your partner is speaking and take notes, if necessary, because you will be asked to summarize what your partner has just said.

(Pass out the prompt card to Candidate A first.)

【Candidate A】 , this is your question.
(Then give a copy to Candidate B.)

【Candidate B】 , this is 【Candidate A's】 question.

<Visual Prompt> (See test paper)

Interlocutor
Now, 【Candidate A】 , please use the next two minutes to prepare your talk.

<Preparation> ☑ (After 2 minutes)

Interlocutor
Now, 【Candidate A】 , please talk for two minutes. Remember I will stop you when your time is up. 【Candidate B】 , please listen very carefully to what 【Candidate A】 says.

Candidate A ☑ (After 2 minutes)

Interlocutor
Now, 【Candidate B】 , you have forty-five seconds to summarize the main points from 【Candidate A's】 presentation.

Candidate B ☑ (After 45 seconds)

Interlocutor
Thank you.

(Pass out the prompt card to Candidate B first.)

【Candidate B】 , this is your question.
(Then give a copy to Candidate A.)

【Candidate A】 , this is 【Candidate B's】 question.

<Visual Prompt> (See test paper)

Interlocutor
Now, 【Candidate B】 , please use the next two minutes to prepare your talk.

<Preparation> ☑ (After 2 minutes)

Interlocutor
Now, 【Candidate B】 , please talk for two minutes. I will stop you when your time is up. 【Candidate A】 , please listen carefully to what 【Candidate B】 says.

Candidate B ☑ (After 2 minutes)

Interlocutor
Now, 【Candidate A】 , you have forty-five seconds to summarize the main points from 【Candidate B's】 presentation.

Candidate A ☑ (After 45 seconds)

Interlocutor
Thank you.

<End of Part III>
This is the end of the test. Goodbye.

<End of the Test>
## Appendix 2: Holistic rating scales

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Utterances are frequently unintelligible and incomprehensible.</td>
<td>• Utterances are intelligible and comprehensible.</td>
<td>• Contributions are both relevant and more than adequate.</td>
<td>• The candidate has no problems with either sounds or utterances.</td>
<td>• The candidate has no problems with either sounds or utterances.</td>
</tr>
<tr>
<td></td>
<td>• Contributions are frequently irrelevant, and inadequate.</td>
<td>• Contributions are relevant, and just about adequate.</td>
<td>• The range, accuracy and appropriateness of language used is adequate.</td>
<td>• Contributions are both relevant and more than adequate.</td>
<td>• The range, accuracy and appropriateness of language used is more than adequate.</td>
</tr>
<tr>
<td></td>
<td>• The range, accuracy and appropriateness of language used is inadequate.</td>
<td>• The candidate generally speaks with ease.</td>
<td>• The candidate speaks coherently in extended turns.</td>
<td>• The candidate speaks fluently with minimal hesitations.</td>
<td>• The candidate speaks fluently with minimal hesitations.</td>
</tr>
<tr>
<td></td>
<td>• The candidate doesn't speak with ease.</td>
<td>• The candidate speaks coherently in extended turns.</td>
<td>• Speech is very well-organized in extended turns.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3: Analytical rating scales

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
<th>Score 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronunciation</td>
<td>• Inaccuracies in individual sounds sometimes prevent communication.</td>
<td>• Comprehensible utterances with easily recognizable sounds, appropriate stress, rhythm, and intonation.</td>
<td>• Inaccuracies and inappropriacies may sometimes occur, but these never prevent communication.</td>
<td>• The candidate is able to produce entirely accurate utterances and sounds with appropriate stress, and natural rhythm, and intonation.</td>
<td>• The candidate is able to produce entirely accurate utterances and sounds with appropriate stress, and natural rhythm, and intonation.</td>
</tr>
<tr>
<td></td>
<td>• Inappropriate use of stress, rhythm, and intonation sometimes prevents communication.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance and Adequacy</td>
<td>• Discourse lacks relevance to the task and contribution is inadequate.</td>
<td>• Discourse is almost entirely relevant to the task and contribution is just about adequate to complete the task.</td>
<td>• The discourse is relevant to the task and the contribution is more than adequate to complete the task.</td>
<td>• The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.</td>
<td>• The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.</td>
</tr>
<tr>
<td>Lexical Use</td>
<td>• The range of vocabulary is inadequate to complete the tasks.</td>
<td>• The range of vocabulary is adequate to complete the tasks.</td>
<td>• There may be some inappropriate use of vocabulary.</td>
<td>• The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.</td>
<td>• The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.</td>
</tr>
<tr>
<td></td>
<td>• Vocabulary is frequently used inappropriately.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical Use</td>
<td>• Too limited a range of structures for task completion.</td>
<td>• An adequate range of structures to complete the tasks.</td>
<td>• There may be some inaccurate structures.</td>
<td>• The candidate is able to use a wide range of structures to complete each task accurately, appropriately and effectively. Errors are rare.</td>
<td>• The candidate is able to use a wide range of structures to complete each task accurately, appropriately and effectively. Errors are rare.</td>
</tr>
<tr>
<td></td>
<td>• Utterances are frequently used inaccurately.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>• Unnecessary hesitation interferes with communication and puts stress on the listener.</td>
<td>• Keeps communication flowing; however, hesitations which do not affect communication may sometimes occur.</td>
<td>• The candidate is able to keep communication flowing smoothly with minimal hesitation.</td>
<td>• The candidate is able to keep communication flowing smoothly with minimal hesitation.</td>
<td>• The candidate is able to keep communication flowing smoothly with minimal hesitation.</td>
</tr>
<tr>
<td>Coherence</td>
<td>• Contribution lacks logical organization.</td>
<td>• Contribution is logically organized in general.</td>
<td>• Contribution is logically organized throughout the task.</td>
<td>• Contribution is logically organized throughout the task.</td>
<td>• Contribution is logically organized throughout the task.</td>
</tr>
</tbody>
</table>
Appendix 4: The survey questionnaire

General English Proficiency Test Advanced Level Pretest
The Questionnaire for the Speaking Examinees

Thank you for taking this speaking test. Please provide your background information in Part I, II, and III, and then answer the questions in Part 4. The questionnaire results shall be used by the LTTC for research purposes only and shall never affect your test results. Please fill out the questionnaire honestly. Thank you for your cooperation.

I. Basic Information
1. Group: ________
2. Test number: ________
3. Gender: ☐ Male ☐ Female
4. Date of birth: Year:_______ Month:_______ Day:_______
5. Mother language: ☐ Mandarin ☐ Taiwanese ☐ Hakka ☐ English
   ☐ Other (Please specify) _______________
6. Have you ever lived in an English-speaking country?
   ☐ Yes Which country? ____________ For how long? year _____ month _____
   ☐ No

II. Occupation/Education
7. Occupation: ☐ Student Name of school _________________ Department _______________
   Year _______________________
   ☐ Non-student Job title__________________ Company _________________
   Years of service____________________
8. Highest education: ☐ PhD ☐ Master ☐ Bachelor ☐ Junior College
   ☐ Senior/Vocational high school ☐ Junior high school ☐ Elementary school
   ☐ Other (Please specify) _______________
9. Highest educational affiliation: ________________________ (Please specify in which country)
   Major ________________________

III. English Learning/Tests Experience
10. When did you start learning English formally? __________ For how many years? __________
11. Have you ever taken a large-scale English test? (e.g. TOEFL, TOEIC, or IELTS)
    ☐ Yes Name of test ______________________ Grades/Scores _______
    ☐ No
12. Have you ever been interviewed in English?
    ☐ Yes Type/Name of the interview__________________ Grades/Scores_______
    ☐ No

When did you answer the following questions?
Year______ Month______ Day______ Hour______ Minute______

S-16
IV. Your opinions about this speaking test

13. Test format
   a. How do you feel about the face-to-face test?
   b. How do you feel about the tape recorder and the video camera? Did this equipment affect your performance? If so, how?
   c. How do you feel about the test room (e.g. the desk, chairs, lighting, etc.)? What arrangements can make you more relaxed and comfortable?

14. The interlocutor
   a. How do you feel about the speech rate of the interlocutor? Did it affect your performance? If so, how?
   b. How do you feel about the interlocutor's attitude (friendly or not)? Did it affect your performance? If so, how?
   c. How do you feel about the interlocutor's instructions (clear or not)? Did it affect your performance? If so, how?

15. Time provided for responding to each part of the test
   a. Was the amount of time provided for responding to Part I (about self-introduction) enough for you? Did you need more or less time?
   b. Was the amount of time provided for responding to Part II (about cell phone) enough for you? Did you need more or less time?
   c. Was the amount of time provided for responding to Part III (about presentation) enough for you? Did you need more or less time?

16. Test performance
   a. On which part do you think you performed best? Why?
   b. On which part do you think you performed worst? Why?
   c. Do you expect to pass this test? Why or why not?

17. Item difficulty
   a. Which part of the test do you consider to be the most difficult? Why?
   b. Which part of the test do you consider to be the easiest? Why?

18. Face validity of the test:
   Do you agree this test can adequately assess your English speaking ability? Why or why not?

   End of the questionnaire.
   Thank you again for your cooperation.

When did you finish the above questions?
Year_____ Month_____ Day_____ Hour_____ Minute_____
Appendix 5: Transcripts of successful examinees' speaking samples

Examiner:X; Candidate A; Candidate B: … means "pauses"

Part I
X: Good morning.
B: Good morning.
X: My name is X, and this is my colleague, D. He is just going to listen to us. And your names are…?
A: I'm A. Nice to meet you.
X: Nice to meet you. And you?
B: I'm B. Nice to meet you.
X: Please.
B: Thank you.
X: First of all, we'd like to know something about you. So I'm going to ask each of you to introduce yourself for one minute. I will stop you when your time is up. Now A, would you please introduce yourself first?
A: OK. Hi, I'm A. I'm 24 years old. Um…I just graduated from UVC last year. I did a double-major program from a bachelor's degree… I specialized in family studies and psychology. Um my family… well there are five people in my family… my parents, me, I have a younger brother and a younger sister. Uh we moved to Vancouver, Canada about ten years ago, and last winter I came back to work in Taipei. Um… my job is related to English teaching… and my hobbies include… watching movies, listening to music, bowling, traveling, and just hanging out with my friends.
X: Um-hm. OK, thank you, A. Now B, could you please introduce yourself?
B: OK. Hi, my name is B, and I've been living in Tainan for 18 years and after that I come here to study in the department of English in National Taiwan Normal University, and I'm currently studying in the Graduate Institute of Translation and Interpretation. Uh there are five people in my family including me… and also my mother, my father, my younger brother, and my great-grandmother… yeah. Uh my brother is now studying also in Shi Da, and specializing in… information technology education. Um… my hobbies are… uh… watching TV, reading novels, and um… playing table tennis. Uh I hope uh after I graduate from the uh… Graduate Institute, I can be a professional translator or interpreter in the future.
X: OK, thank you, B.
B: Thank you.
X: Now, I'm going to ask each of you a few questions about yourselves. B…
B: Um-hm.
X: When you were a child, what kinds of books did you like to read?
B: Um… I liked to read books on fairy tales, or on some… I also liked to read some sci-fi… uh science… science fiction. And um… but mostly I… I would like to read some… um… interesting novels like um… um… yeah, including what I just mentioned uh science fictions and uh… fairy tales or uh… things about ghost stories.
X: OK, thank you. How about you, A? When were you a child, what kinds of books did you like to read?
A: Well uh… I liked to read… Chinese folk tales or Japanese folk tales… fairy tales… and to be honest… some comic books. Um… that's… but I… I never have any interest in reading ghost stories, that kind of thing.
X: Indistinct. And A, have you ever read a book more than once?
A: Oh, yes.
X: Why?
A: Um… well… I think… because… I really enjoy the story… um that's the main reason for reading a book more than once… or maybe sometimes I just have a bad memory, so reading the book a second time is kind of… like reading it for the first time for me… sometimes.
X: OK, thank you. How about you, B? Have you ever read a book more than once?
B: Yeah, sure.
X: Why?
B: Because uh… most of them are… comic books, and I read them more than once because uh… I think that doesn't take a lot of mental functioning. I mean you don't have to uh… think… a lot on its contents, so… you just relax and reading it.
X: OK, that's fine. And B, are you planning to read any particular author's books in the near future?
B: I think I will, 'cause uh, it's a need of my… uh… of what I learn now, 'cause I'm….
X: And whose books are they?
B: Uh, I think I would um…read some…poetic works…yeah…by some particular poets.
X: OK, thank you. How about you, A? Are you planning to read any particular author's books in the near future?
A: Actually, I'm just planning to read uh…Helen Fielding's? Actually, I don't quite remember her last name, but it's the author who wrote Bridget Jones' Diary. I've heard that it's a very interesting novel, and I have just recently watched the movie, so…And maybe um…J.K. Rowling's future Harry Potter…um books.

Part II
X: OK, thank you. Now, you two are going to work together to complete a task. A…. Please do not show your card to your partner. …. Here's the situation. Each of you is going to buy a cell phone for yourself. A, you have information on three cell phones. And B, you have different information on the same three cell phones. You will first need to exchange information to fill the chart. You may write on the card. You will have two minutes to exchange information. I will stop you when your time is up. Now please begin, and A, please speak first.
A: Uh, OK. Uh, I have some information on each of the cell phones.
B: Um-hm.
A: Uh, cell phone A costs…four thousand, three hundred NT dollars.
B: Um-hm.
A: And cell phone B costs eight thousand, eight hundred…eight hundred and eighty dollars.
B: Um-hm.
A: Uh, the battery duration for cell phone C is one hundred and thirty hours.
B: Um-hm.
A: Cell phone C doesn't have…Internet…um access…and it doesn't have voice-dialing function.
B: Um-hm.
A: It's not brand-new, and therefore it is…second-handed.
B: OK.
A: Yeah, and…something else…for cell phone B, it has Internet access, um…and…whereas for cell phone A, it doesn't have built-in games….
B: Um-hm.
A: But it is brand-new.
B: Um-hm. … OK. … And um, there's something about…the…the rest of the information. On cell phone C, the price is two thousand, seven hundred dollars, and its battery duration is one hundred and thirty hours. It is uh…I'm sorry uh…about cell phone A, the battery duration is one hundred and fifty hours, and cell phone B is…is the same.
A: OK.
B: And uh…concerning Internet access, cell phone A doesn't have it.
A: OK.
B: Um-hm. Uh…and cell phone B doesn't have any built-in games, but cell phone C does. And voice-dialing, uh cell phone B…has this function but cell phone A doesn't. And cell phone B is a brand new…cell phone.
A: OK.
X: OK? Thank you. Now during the next two minutes, you and your partner are going to discuss both the advantages and disadvantages of each phone. Remember, this is a discussion. Focus your talk on the advantages and disadvantages of each cell phone. I will stop you when your time is up.
B: Um-hm.
X: Now please begin the discussion, and B, please speak first.
B: OK. Hmm…A, I think um…I should buy a cell phone. Can you make an analysis of the advantages and disadvantages of the three ones for me?
A: Uh, OK. I think actually, cell phone B is very expensive.
B: Um-hm.
A: Uh eight…almost nine thousand dollars.
B: Yeah.
A: Uh…but that's probably because it's brand-new….
B: Um-hm.
A: Um, if you insist on buying a brand-new product, then…it has voice-dialing function and it has…can…uh it has Internet um…access, too.
B: Um-hm.
A: So well…it depends on your budget, but…mmm…
B: But I…I think I can afford only…cell phone A and cell phone C 'cause I have only like five
thousand dollars, so... so now I'm concerning... so now I'm considering um... the functions of the
two, so... yeah... which one do you think is better?
A: Well, I... well the advantage for cell phone C is that it's the cheapest...
B: Um-hm.
A: ... among the three....
B: Yeah.
A: And personally, I don't mind using a second-hand um... product.
B: Um-hm.
A: Um... and... I... don't think... having Internet access is that important to me.
B: Um-hm.
A: Yeah.
X: And what do you think, B?
B: Um... well, I think for me, um... yeah. I wouldn't mind using a second-hand cell phone, but I
think uh... um... a brand-new phone would be... like... make me feel assured, like... its....
A: Safer....
B: Gonna be... yeah... gonna be... lasting longer, and... I'm not so um... interested in games in a cell
phone, so....
A: So cell phone A seems to be the best choice.
B: Yeah.

X: Um-hm. Thank you. Now, B, in thirty seconds, please tell us which cell phone you wish to buy,
and why.
B: OK. Uh... according to my budget and my uh... interest, I think I would buy cell phone A 'cause
it's... not so... expensive and it has... basic functions. It has um... long battery duration, and most
important, it is brand-new. So I think I will choose cell phone A.
X: Thank you. Now, A, in thirty seconds, please tell us which cell phone you will buy, and why.
A: OK. I think I'll choose cell phone A as well um... since its... it's price range is closer to my budget,
and... also I'm not that interested in the Internet access or built-in games... uh in a cell phone.
Um... like B said, a brand-new product will probably... have um... comes with a guarantee... with
the... the cell phone itself....
X: OK, thank you. Now, A, please tell us, are cell phones dangerous to use while driving, and why?
A: While driving.... Uh... yes, I believe it is dangerous to use cell phones uh... while you're driving.
The most obvious reason being it's... you... a person might be easily distracted by... uh... the cell
phone – the ringing of the cell phone or the fact that she has to talk... um... on the cell phone while
driving. So... I wouldn't recommend it. In fact, I think there should be laws set up... um... against this
kind of action.
X: Um-hm.
A: And... um... actually, what am I supposed... are there... a second part of the question? ... No. OK,
that's it.
X: OK, thank you. Now, B?
B: Um-hm.
X: Please tell us, why have cell phones become so popular among young people? You have one
minute to answer. I will stop you.
B: OK. I think in the past um... people are... don't... feel very... it's not very convenient for people
to... well... get in touch with each other 'cause there's no... there's nothing like a cell phone... which
is very, very um... mobile, and people can use it to... get in touch with people anywhere, so I think
that's the major reason why it becomes so popular, especially among young people, 'cause young
people need to, like, contact others very often, when they are indistinct, or they are, like, to go
party, to go... to go out playing, or something. And um... and also, cell phones are not a kind of
product which is very, very expensive... for the... young people. So I think that's why.

Part III
X: Thank you, B. Now... thank you. ... In a few seconds, I will give each of you a card with a
question on it. Each of you will have two minutes to think about what you are going to say. Then
you will have two minutes to present your talk to your partner. Listen carefully while your partner
is speaking and take notes if necessary, because you will be asked to summarize what your partner
has just said. Now, A... this is your question.
A: OK. Thank you.
X: B, this is A's question.
B: OK.
X: OK? Now, A, please use the next two minutes to prepare your talk.
After two minutes.
X: Now, A, please talk for two minutes. Remember, I will stop you when your time is up.
A: OK.
X: B, please listen very carefully to what A says.
B: OK.
A: OK, um…we all know that nowadays um…a lot of scientists, they are working very hard on finding cures for um…many different types of serious illnesses. And one way to reach their…one way to...come up with some findings that um…that are useful is to use laboratory…uh animals in laboratory tests. Um…I…personally, I think it is ethical to use animals in laboratory tests. Uh, but the most important thing…the most important point being that these…laboratory experiments should be well-designed, and well-regulated. And these um…animals that um…these laboratory animals uh be kept and raised and maintained in a humane way…and even if they have to…sacrifice their lives for us human…um…it should be done in…like, like I said, in a humane way…not just…ruthlessly…you know. And um…um…have some background in um…science…scientific fields, and we used to um…raise and perform laboratory tests on laboratory…on these animals…like rats, or monkeys. Um…but we…with well-designed experiment, it is really helpful, and um…in our reports…um…of course we will…give full credit to these animals. And um…also the people who designed um…who designed these experiments. Um…
X: OK. Thank you, A.
A: Oh, OK. Thank you.
X: Now, B, you have 45 seconds to summarize the main points of A’s presentation.
B: OK. Um, basically, A is for um, the proposal of um…using animals in laboratory tests. But um…she…presents some um…prerequisites. The first one is um…the environment of the laboratory should be well-designed and well-regulated. And the other is, the animals used should be kept and maintained, or dealt with in a humane way. And she also used some of her personal experiences to back up the proposal.
X: Um-hm. And…?
B: And I think, so…so uh…basically, yeah…the uh…she think it…it is ethical to use animals in laboratory tests.
X: Thank you, B.
B: Thank you.
X: OK. Now…OK, so…B, this is your question. …A, this is B’s question.
A: OK.
X: OK? Now, B, please use the next two minutes to prepare your talk. Thank you.
B: OK.

(After two minutes.)
X: Now, B, please talk for two minutes. I will stop you when your time is up. And A, please listen carefully to what B says.
A: OK.
B: Basically, I think Taiwan should build…should build more nuclear power plants. But under the major premise…that is…um…the government should do some precautionary measures…to prevent um…possible danger or harm to the…citizens in Taiwan. And…why? Why should Taiwan build more nuclear power plants? I think the…the most important reason is that…we are in great need of power…of electrical power. Um…for one thing, uh the factories, the industries is keeping growing, so I think uh…the factories need much power…nuclear…uh much…electrical power. And also, the people, the citizens…the citizens in Taiwan…are um…needing…needing uh electrical power…indistinct more…uh they… According to some statistics, um…the amount of electrical power used every…each year is increasing. So based on the main…based on the recent…I think Taiwan should build more nuclear power plants to provide electrical…electricity to the citizens in Taiwan. And also, one thing we should not ignore is that um there’s…still not…there's still no um…substitute energy form…yeah…provided now. So…uh…in supply of um…the need for the people in Taiwan…uh I think…I think it is um…necessary…yeah.
X: OK. Thank you, B. Now, A, you have 45 seconds to summarize the main points from B’s presentation.
A: OK. B believes that Taiwan uh…needs to build more nuclear power plants. Um…but um…the most important thing being the government…should um…the government should uh…should have some precautionous measures to minimize the potential dangers…of building…more nuclear power plants. And…the reason we need more nuclear power plants, like B has said, is that in fact, in reality, we need more um…electrical power. The…electricity consumption is increasing in
Taiwan…and also, currently, there aren't any suitable alternatives for power supply.

X: Thank you, A. OK, thank you. This is the end of the test. Thank you.
A: Thank you.

EXAMINER'S COMMENT For CANDIDATE A

Pass 5
The candidate has entirely accurate pronunciation in both individual sounds and longer utterances. She responds relevantly to all questions, and is able to elaborate on all the main points she makes. She demonstrates flexible and appropriate use of vocabulary and uses a wide range of syntactic structures very accurately. She is able to communicate fluently without any unnatural hesitation. Her contributions to all tasks is well organized.

EXAMINER'S COMMENT For CANDIDATE B

Pass 4
The candidate produces easily comprehensible and accurate utterances. He responds to all questions relevantly, and is able to elaborate on most of his main points. He is able to use appropriate vocabulary to express himself and a good range of syntactic structures; however, occasional errors which do not affect communication still occur. Generally, the candidate is able to communicate fluently with only occasional, unnatural hesitations. In addition, the candidate's contributions to all tasks are generally well organized.
Bibliography

語言訓練測驗中心. 2001. 《全民英語能力分級檢定測驗九十年中級測驗初、複試成績統計報告》台北: 語言訓練測驗中心.

語言訓練測驗中心. 2001. 《全民英語能力分級檢定測驗九十年第一次中高級測驗初、複試成績統計報告》台北: 語言訓練測驗中心.

語言訓練測驗中心. 2001. 《全民英語能力分級檢定測驗九十年第二次中高級測驗初、複試成績統計報告》台北: 語言訓練測驗中心.


語言訓練測驗中心. 2000. 《全民英語能力分級檢定測驗八十九年第一次中高級測驗初、複試成績統計報告》台北: 語言訓練測驗中心.


鄧慧君. 1998. ＜文類及題型對英語聽力測驗之影響＞，《英語教學》，1998，第 23 卷，頁 5-18。


Berne, J. E. 1992. The Effects of Test Type, Assessment Task, and Target Language Experience on Foreign Language Learners’ Performance on Listening Comprehension Tests. Diss, University of Illinois at Urbana-Champaign.


Skehan, P. & Foster, P. 1997. ‘Task type and task processing conditions as influences on foreign language performance’ Language Testing Research 1(3): 185-211.


