A Content Validity Study of TEM-8 Reading Comprehension (2008 -2010)

Bijin Dong
School of Teacher Education
Level IV English, Spring 2011
School of Teacher Education
Kristianstad University
Tutor: Jane Mattisson
Table of Contents

Abstract ........................................................................................................................................................................ 1

1. Introduction .................................................................................................................................................................... 2

  1.1 Aim........................................................................................................................................................................... 3

  1.2 Material .................................................................................................................................................................... 3
  1.2.1 The 2000 Teaching Syllabus for Reading Comprehension of TEM-8 ............................................................ 4
  1.2.2 The 2004 Test Syllabus for Reading Comprehension of TEM-8 ................................................................. 4

  1.3 Method..................................................................................................................................................................... 6

2. Theoretical Background / Previous Research ........................................................................................................... 9

  2.1 Definitions of Key Concepts ................................................................................................................................ 9
  2.1.1 Validity ............................................................................................................................................................... 9
  2.1.2 Content validity ............................................................................................................................................... 11
  2.1.3 TEM and TEM-8 ............................................................................................................................................... 12
  2.1.3.1 Settings and Rubrics of TEM-8 .................................................................................................................. 13
  2.1.3.2 Rubrics of the Reading Comprehension Component of TEM-8 ............................................................... 13
  2.1.3.3 Test Components ....................................................................................................................................... 14

  2.2 Theoretical Concerns in Communicative Language Testing ................................................................................. 16
  2.2.1 Bachman’s View of Communicative Language Ability .................................................................................. 16
  2.2.2 Correspondence Between Test Performance and Language Use .................................................................... 17
  2.2.3 Framework of Task Characteristics .............................................................................................................. 18

  2.3 Previous Research .................................................................................................................................................. 22
  2.3.1 Previous Research on Validity .......................................................................................................................... 22
  2.3.2 Previous Research on TEM-8 .......................................................................................................................... 23
  2.3.3 Previous Research on Reading Comprehension Test (RCT) ........................................................................ 24

3 Analysis and Discussion ........................................................................................................................................ 25

  3.1 Input ....................................................................................................................................................................... 25
  3.1.1 Total Length ...................................................................................................................................................... 25
  3.1.2 Difficulty Variables ....................................................................................................................................... 26
  3.1.2.1 Average Sentence Length ......................................................................................................................... 26
  3.1.2.2 Distribution of New Information ............................................................................................................. 28
  3.1.2.3 Readability ............................................................................................................................................... 29
  3.1.2.4 Sentence Types ....................................................................................................................................... 31
  3.1.3 Topics ............................................................................................................................................................... 33
  3.1.4 Genres .............................................................................................................................................................. 34

  3.2 Expected Response ............................................................................................................................................... 35

4. Summary and Conclusion ......................................................................................................................................... 40

  4.1 Major Findings ...................................................................................................................................................... 40

  4.2 Implications ......................................................................................................................................................... 41

  4.3 Limitations .......................................................................................................................................................... 42

References .................................................................................................................................................................... i

Appendices .................................................................................................................................................................... iv
Appendix 1 .................................................................................................................................................................... iv
Abstract

TEM-8 (Test for English Majors, Band 8) is a large-scale nationwide standard test for English majors in China. It has been increasingly recognized in China and has already become one of the criteria for judging English teaching and learning in China. It is necessary to pay more attention to its development. Meanwhile, reading comprehension is regarded as vitally important in acquiring a language. It is worthy investigating how to measure reading comprehension ability accurately and properly in language testing and whether reading comprehension in TEM-8 can reflect the test-taker’s reading ability. This thesis analyzes the content validity of the reading comprehension part of TEM-8 from 2008 to 2010 in terms of the new framework of task characteristics revised from Bachman and Palmer’s framework (1996: 49), and in accordance with The Teaching Syllabus for TEM-8 and The 2004 Test Syllabus for TEM-8. This study has shown that TEM-8 reading comprehension part has relatively high content validity, basically corresponding to the requirements in the teaching syllabus and test syllabus. However, there are some deficiencies influencing the content validity of TEM-8, such as limited test method, limited data recources and time.

Key words: TEM-8, Reading Comprehension, Content validity
1. Introduction

As a necessary skill in language learning, reading is critical in English language teaching and learning. Reading competence-training is regarded as a major goal of English teaching in the syllabi of English teaching in secondary and colleges (both for English majors and non-English majors). Thus, reading takes a large proportion in large-scaled language tests. Accordingly, designing a valid, reliable and practical comprehension testing and ensuring its validity is crucially important in foreign language testing. In recent years, language teachers and scholars on testing have paid more attention to improve the quality of reading tests (Bachman1996 & Alderson 2000 & Jin Yan 2002).

TEM-4 and TEM-8 (Test for English Majors Band 8), has been increasingly recognized in China and over the world in recent years as the only national standard tests for English majors, such as ACCA in Britain (It requires students to provide the TEM certificate to prove their English proficiency). TEM-4 is for learners at the foundation stage and TEM-8 is for those at the end of a four-year undergraduate programme which has been implementd for many years since 1900 and they also have been employed as the criteria for judging English teaching and learning in China. What’s more, there are only a few studies having been done on the TEM-8. Hence, it is necessary to pay more attention to the TEM’s development.

At the same time, reading comprehension is always used as the main means of acquiring a language. In China, second language reading ability of English is in great demand when the language of science and technology and advanced research as well as a global language.

On account of the significance of TEM-8 and the important role of reading comprehension ability, it is requisite to raise concerns on the evaluation and development of effective tests of reading comprehension. The focus of the present study is on reading comprehension part of TEM-8.

Furthermore, language testing establishes the target ability by measuring the behavior or manifestation of this ability (Li Xiaoju 1997: 30). That is to say that it is the behavior of target ability which is measured in a given test. Nevertheless, it is impossible to measure all the behavior of target ability in one test. Consequently, it is hard and important to define the sample behaviour based on the test syllabi. This paper will examinationine “the extent to
which the tasks required in the test adequately represent the behavioral domain in question” (Bachman 1990:245), i.e. the content validity.

This paper focuses on exploring the content validity of reading comprehension task in three papers (2008-2010) by making a detailed comparison between the TEM-8 test content and test syllabus and teaching syllabus for English majors in terms of the newly revised one based on Bachman and Palmer’s framework for validating content of reading test.

1.1 Aim

This thesis aims to explore the content validity of the reading comprehension part of TEM-8 from 2008 to 2010 in terms of the new framework of task characteristics revised from Bachman and Palmer’s framework (1996: 49).

There are two purposes in this research. The first is to explore the content validity of reading comprehension part of TEM-8 from 2008 to 2010 in terms of the characteristics of input mainly based on The 2000 Teaching Syllabus for TEM-8 and The 2004 Test Syllabus for TEM-8. The characteristics of input consisting of total length, difficulty variables, topics and genres are investigated. The second is to analyze the content validity of reading comprehension of TEM-8 from 2008-2010 in terms of characteristics of expected response based on these two syllabi. The expected response includes the coverage of reading skills.

1.2 Material

The materials employed in this thesis are The 2000 Teaching Syllabus for TEM-8 for Reading Comprehension, 2004 Test syllabus for TEM-8 for Reading Comprehension, and three examination papers of TEM-8 from 2008-2010.

Alderson (2000) states that “A test’s specification provides the official statement about what the test tests and how it tests it” (Alderson et al. 1995:9). However, it is not easy to get test specification since it is only for internal staff and it is confidential while syllabus is a public and simplified document which is for “teachers and students who wish to prepare for the test, to people who need to make decisions on the basis of test scores, and to publishers who which to produce materials related to the test (Alderson et al. 1995:9). As this thesis assesses the
content validity of the TEM-8 reading comprehension parts, and TEM papers are designed to check whether student master the abilities required in the teaching syllabus, so one of the criterions should be the teaching syllabus. Meanwhile, the validity of language tests should concern about a detailed description of both the abilities to be measured in detail and the facets of the test methods. The testing syllabus describes the abilities to be measured in detail and they are drawn up on the basis of teaching syllabus. Therefore, testing syllabus is needed in this study as well.

1.2.1 The 2000 Teaching Syllabus for Reading Comprehension of TEM-8

As the guidelines for educational decision-makers and teachers, the English language teaching syllabus is supported by the Ministry of Education in China. It has expounded the training objectives, curriculum arrangement, teaching requirements, teaching principles, tests and assessment. Some specifications concerning reading comprehension of TEM-8 are extracted from the teaching syllabus (Higher Education Institution English Teaching Committee Group is abbreviated as Higher Education Institution in the following, 2000: 5) and described below:

1) Be able to comprehend the editorials and book reviews in newspapers and magazines in America or Britain, and historical biographies and literary works to a certain degree of difficulty, which are published by English speaking countries.
2) Be able to analyze the viewpoints, discourse structures, language characteristics and figurative devices of the subjects above.
3) Be able to read at the speed of 140-180 words per minute.

1.2.2 The 2004 Test Syllabus for Reading Comprehension of TEM-8

The revision of Test Syllabus (Higher Education Revision Group of the 2004 Test Syllabus for TEM-8 abbreviated as Higher Education Revision Group in the following 2004: 3) for TEM-8 was finished in 2004. Some specifications related to the reading comprehension of TEM-8 are cited as follows.

(a) Test requirements
(1) To be able to comprehend the editorials and book reviews in newspapers and magazines in America or Britain.
(2) To be able to comprehend historical biographies and literary works to a certain degree of difficulty.
(3) To be able to grasp the main ideas, distinguish the important facts or details, to understand both the explicit and implicit meanings, to make correct deductions and inferences, and to analyze writer’s intention or attitude to audience and to topic, discourse structures, language characteristics and figurative devices.
(1) To be able to duly modify reading speed and apply reading skills.
(2) 20 multiple choice questions items to be finished within 30 minutes.

(b) Test formats
There are several (normally four) passages with approximately 3000 words in this part. After each passage, there are several multiple choice questions with 20 multiple choice questions in all. Test-takers are asked to choose the most appropriate answer from the four choices.

(c) Test purposes
(1) To test the testees’ ability to obtain information through intensive reading.
(2) To test the extent to which testees apply their reading skills and strategies.
(3) Besides, both accuracy and speed at about 150 words per minute are expected in this part.

(d) Principles of material selections
(1) A wide range of topics including society, science and technology, culture, economy, daily life, biography and so on.
(2) A variety to text genres such as: narration, description, exposition, argumentation, advertisements, instruction books, and charts and graphs.
(3) Key words which may affect the global comprehension of the text should be substantially within the requirements of the syllabus.

TEM-8 Papers from 2008 to 2010 are designed by the 2004 test syllabus. The format of TEM-8 is shown in the table below.

Table 1: The format of TEM-8

<table>
<thead>
<tr>
<th>COMPONENT</th>
<th>NO. OF ITEMS</th>
<th>POINTS</th>
<th>WEIGHTING(%)</th>
<th>TIME(mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part One Listening Comprehension</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>a) mini-lecture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) conversation or interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) news broadcast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part Two Reading Comprehension</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>Part Three General Knowledge</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Part Four Proof Reading &amp; Error</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Correction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since the most important quality of test interpretation or use is validity and one of the first characteristics of a test is to examine its content (Bachman 1990:244), the papers and syllabi from the three years are the main source of data. Since 2008 to 2010 TEM-8 papers are the latest versions available and have not been the object of critical attention, they have been selected for the purposes of the present study, so this thesis chooses them as the research subject. The authentic version of the TEM-8 papers from 2002 to 2010 edited by Li (2010) was chosen for analysis. In order to ensure the accuracy of the content, another volume of the TEM-papers (Sheng 2010) are used to check it. The 2000 Teaching Syllabus for TEM-8 and The 2004 Test Syllabus for TEM-8 published by Shanghai Foreign Language Education Press are also collected for this study.

1.3 Method

The research project is divided into three stages. First, a framework of communicative reading assessment is constructed, consisting of framework of the communicative test characteristics and the nature of reading. Second, by analyzing the validity of the reading comprehension part of TEM-8 papers from 2008 to 2010 in terms of The 2000 Teaching Syllabus for TEM-8 and The 2004 Test Syllabus for TEM-8, some insights are gained in the following analysis. Third, conclusions are reached and proposals are brought forward to improve the existing test in order to guarantee that future tests will better reflect the English language performance of the test takers. According to the stages identified above, the research questions are presented as follows:

Question 1. What are the critical characteristics of reading tasks?

In order for the inferences about test takers’ ability to use language in the target language use (TLU) domain to be made, that is, for the test validation, a language test should consist of language use tasks. Bachman and Palmer (19996: 47-57) provide a much more detailed
framework of language characteristics build on that proposed by Bachman in 1990. However, according to Bachman and Palmer(1996: 47), “test developers need to make some modifications in the specific characteristics…will not necessarily follow the order described in the framework”. For the sake of making the assessment more practical, some modifications are made to set a new framework on the basis of the research of Bachman and Palmer and characteristics of Reading Comprehension of TEM-8. The new framework is regarded as a task characteristic checklist for comparing characteristics of target use tasks and test tasks in TEM-8, which serves to analyze the content validity. Below is the new framework based on the original framework of Bachman and Palmer (1996:49-50).

Table 2: The framework of language characteristics

<table>
<thead>
<tr>
<th>Task characteristics</th>
<th>Characteristics of the setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physical characteristics</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
</tr>
<tr>
<td></td>
<td>Time of task</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristics of the test rubrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
</tr>
<tr>
<td>Structure</td>
</tr>
<tr>
<td>Time allotment</td>
</tr>
<tr>
<td>Scoring method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristics of input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length</td>
</tr>
<tr>
<td>Difficulty variables</td>
</tr>
<tr>
<td>Average sentence length</td>
</tr>
<tr>
<td>Distribution of new information</td>
</tr>
<tr>
<td>Facility value &amp; Readability</td>
</tr>
<tr>
<td>Sentence types</td>
</tr>
<tr>
<td>Topics</td>
</tr>
<tr>
<td>Genre</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic of expected response</th>
</tr>
</thead>
<tbody>
<tr>
<td>The coverage of reading skills</td>
</tr>
</tbody>
</table>

The reasons for the modification of the original one are explained here: to start with, Reading Comprehension of TEM8 is the only part to be analyzed in this thesis. Thus, some characteristics are not needed to analyze while some characteristics concerning to the nature of reading and the requirement of syllabi such as total length, difficulty variables, topics and genres are requisite to be illustrated directly in the new framework. The average sentence
length, distribution of new information, readability and facility value and sentence types are
the variables to influence the difficulty of reading. However, textual, functional and
sociolinguistic characteristics are not discussed in this thesis because of time and energy
limitation. Second, since Reading Comprehension of TEM-8 only adopts the selected
response type, many irrelevant factors listed in the original framework are removed in the new
framework. Third, the fifth aspect listed in the original framework, the relationship between
input and response is removed as a result of impossibility of accessing to the test-takers’
performance.

Question 2. What are the specific characteristics of the TEM-8 reading comprehension section
in recent three years?

The test papers of the three years have undergone through and detailed examinationination
and assessment. The qualitative and quantitative studies have been implemented, which
consist of three steps: counting, calculation and comparison.

Counting
The total length, average sentence length, distribution of new information, sentence type,
distribution of genres, topics and the skill coverage of the data are to be analyzed by counting.

Calculation
Calculation of the readability: The formula \( RE = 206.84 - 0.85wl - 1.02sl \) to assess how easily
written materials can be read and understood. Here \( wl \) represents average number of syllables
in 100 words, \( sl \) refers to the average number of words per sentence.

Comparison
As the content of TEM-8 is written based on syllabi, comparisons are used to scrutinize
whether the empirical study results concerning the characteristic of input and expected
response of reading comprehension in three examinationination papers are in consistent with
the requirements of the syllabi. In addition, comparison is also employed to examinationine
the internal consistency of reading comprehension of these three examinationination papers to
each other.

Question 3. What is the content validity of the TEM-8 reading component?
Owing to the limited materials and time resources available to the study, this paper only aims to explore on the aspect of content validity. According to the validity of the TEM-8 reading by way of content analysis, comparisons are made between the results of the survey and the requirements of the teaching syllabus (2000) and testing syllabus (2004). Finally, the merits and demerits are presented to provide proposes and recommendations for the future improvement of TEM-8.

2. Theoretical Background / Previous Research

This section comprises of three sections. The first section presents key terms related to the study. Section two provides theoretical assumptions on which this paper is based on. Section three reviews the previous studies of validity, TEM-8 and the Reading Comprehension of test.

2.1 Definitions of Key Concepts

This section discusses some key terms in this study, namely validity, content validity, TEM and TEM-8, and the Reading Comprehension of Test.

2.1.1 Validity

The primary concern when designing and developing a language test is its usefulness. Hence, the usefulness of a test is considered to be the most essential quality. Bachman and Palmer (1996:18) offer a formula that test usefulness can be described as a function of several qualities: Usefulness = Reliability + Construct validity + Authenticity + Interactiveness + Impact + Practicality. According to Bachman and Palmer (1996), here reliability refers to the consistency of measurement. When construct validity is defined as the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores. Meanwhile, authenticity pertains to the degree of correspondence of the characteristics of a given language test task to the features of TLU task while interactiveness is defined as the extent and type of involvement of the test taker’s individual characteristics in accomplishing a test task. Another quality of tests is their impact on society and educational systems and upon the individuals within those systems. The last test quality, practicality pertains to primarily to the ways in which the test will be employed and whether it will be developed and used at all.
Bachman and Palmer point out, although it is the overall usefulness of the test that is to be maximized rather than the individual qualities that affect usefulness, among these six qualities, reliability and validity are most essential for a language test (1996:18).

The reliability of a language test is quite easy to be showed by statistical analysis. As the concept of reliability is from general metrology, it is relatively stable while the validity of test from a socio-scientific metrological concept, is much more sophisticated than reliability.

Validity refers to “the degree to which that evidence supports the inferences that are made from the scores” (American Psychological Association 1985:9). It is the inferences regarding specific uses of a test are validated not test itself. Validity as a theoretical rationale support the adequacy and appropriateness of inferences and actions based on test scores (Messick 1989:13).

The evidence collected in support of a particular test use can be grouped into four types: construct validity, content validity, criterion validity and face validity. Construct validity is concerned with the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs (Bachman 1990:225). Construct validity is indeed the unifying concept that integrates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationships (Messick 1980:1015). As seen from the definition above, it is concerned how performance on tests is in accordance with predictions that we make based on a theory of abilities, or constructs. Thus, the principal task of the study on construct validity is to make clear the ability to be tested and to find out the theory employed in the language testing. Content validity is established by specifying the domain to be sampled for testing and then selecting test items to present that domain (Davies et al. 1999:222). In other words, content validity depends on a careful analysis of the language being tested and of the particular course objectives. So content validity is the most essential part of the test assessment. Criterion validity consisting of concurrent validity and predictive validity are set up statistically according to the closeness of a test to its criterion; there may, for instance, be an existing test or some other measure within the same domain (concurrent domain) or a future test or other measure (predictive validity) (Davies et al. 1999:222). Face validity refers to the degree to which a test appears to measure the knowledge or abilities it claims to measure, as judged by
an untrained observer (such as candidate taking the test or the institution which plans to administer it) (Davies et al. 1999:69).

As seen from above, validity of a test should be supported by evidence. Some frameworks can be employed for evidence-gathering, such as the Bachman and Palmer’s framework which will be discussed in the next section. As this paper intends to study the content validity of TEM-8 reading comprehension parts, it is necessary to review content validity in detail.

2.1.2 Content validity

As Bachman (1990) claims, one of the first characteristics of a test that a prospective test user examinations is its content. The test designers begin with a definition of the content or ability domain from which they come up with items and test tasks in developing a test. Therefore, the consideration of test content is thus an important part of both test development and test use, and an analysis of the content of a test is a essential part of validity as well (Bachman 1990:244). Thus, this paper aims to explore the content validity of TEM-8 reading comprehension section. Content validity is the representative and sampling adequacy of the content- the substance, the matter, the topics of a measuring instrument (Alderson et al. 1995:173). A common way to evaluate the content validity is to analyze the content of a test and to compare it with a statement of what the content should be. “Such a content statement may be the test’s specifications, it may be a formal teaching syllabus or curriculum, or it may be a domain specification” (Alderson et al. 1995:173). Henning (1987:94) says, “it is precisely in the area of content validity of achievement measures that criterion- or domain-referenced tests have certain profound advantages over more traditional norm-referenced tests.” Huges (1989:22) states that “the greater a test’s content validity, the more likely it is to be an accurate measure of what it is supposed to measure.” There are two aspects of content validity in Bachman’s (1990) opinion: content relevance and content coverage.

The investigation of content relevance requires the specification of the behavioral domain in question and attendant specification the task or test domain (Messick 1980: 1017). So it involves the specification of the ability domain and the specification of test method facets. The following are the elements to specify the ability domain and test method facets: (1) what it is that the test measures; (2) the attributes of the stimuli that will be presented to the test taker; (3) the nature of responses that the test taker is expected to make (Hambleton 1984).
That is to say, the specified language ability, population and possible tasks can be defined in this process. The content coverage is the second aspect of the content validity, which assures the tasks required by the test representatives of that domain. Hence, the more representative and typical the task is, the more valid it is thought to be. An example could be followed, drawing multiple samples of tasks from the domain, to determine the extent to which different sets of tasks are equivalent, as a way of demonstrating content coverage (Bachman 1990:245).

Content validity is significant for both language teaching and language testing. If the content is not validly selected, then score interpretations are liable to be distorted (Wood 2001:148). Hughes (1989:27) states that the greater a test’s content validity is, the more likely it is to accurately measure what is supposed to measure. He (2000) also proposes that if a test in which some areas are not tested is likely to become areas ignored in teaching and learning, it is likely to have a detrimental backwash (Which Hughes (1989:1) defines as “the effect of testing on teaching and learning) on teaching and learning. If only some contents are covered, it will produce unfavorable guidance to the teaching and improving the practical language abilities.

2.1.3 TEM and TEM-8

TEM has been implemented nationwide in China for nearly twenty years, whose score has been increasingly used as measures of proficiency in English as a foreign language throughout China. The participants of the test has been increasing every year. According to the official figures, up to 2009, there are more than 280000 people having attended the test of TEM-4 while there are more than 180000 people for TEM-8. The increasing number of test takers stimulates a tendency in China that a great number of individual’s careers or education decisions are considerably affected, such as applying for admission to an educational program or seeking employment as well as advancement in a career.

TEM is designed to check students’ English language abilities according to the relevant teaching syllabi. The two national English language teaching (ELT) syllabi for English language majors were drawn up under the auspices of the State Education Commission of China (SECC). They were published in 1989 and 1990 respectively. One of them is for the foundation stage, i.e. the first two years of a degree programme in English language and literature; the other is for the third and fourth years of the same programme. “Their primary
purpose is to serve as guidelines for educational decision-makers and teachers alike when it comes to the development of ELT curriculum, ELT programmes and ELT materials in various universities across China” (Zou 1997:1). To emphasize the continuity of English learning at the foundation and higher stages, the two national ELT syllabi were combined into one and published in 2000. The new syllabus not only includes the different requirements and characteristics in the foundation stage and higher stage, but also emphasizes the importance of learning English basic ability. TEM is a syllabus-based achievement or criterion-reference proficiency test in nature, that is, a test administered to measure the effects of a language instruction program in accordance with the requirements of a certain syllabus (Huges 1989).

2.1.3.1 Settings and Rubrics of TEM-8

TEM is not an oral test but a written examination in which each test-taker will receive a test paper and an answer sheet. TEM-8 consists of six parts in which the scoring of listening comprehension, reading comprehension, and general knowledge will be done by machine while proof-reading & error correction, translation and writing will be scored manually.

TEM-8 begins in the morning and participants have to complete the examination in 195 minutes. What’s more, the humidity, temperature and draught should be under control at the testing time. Participants are also required to prepare pencil, pen, eraser and so on by themselves. Students of the same class usually sit together and invigilators are usually their familiar teacher to reduce students’ tension.

2.1.3.2 Rubrics of the Reading Comprehension Component of TEM-8

Reading comprehension in the TEM-8 examination contains 4 passages for which there are several multiple choice questions and the number may vary from 4 to 6 within the total number of 20. All these items should be completed within 30 minutes. In the beginning of this part, a brief direction will be given to test-takers in the target language through written words in the paper. Test-takers are informed to choose the most appropriate answer from the four choices. In order to guarantee the high scoring reliability of the test, the scoring is mechanical, which ensures the full objectivity without any subjective opinions on the answers of test graders. In addition, mechanical marking is much simpler, rapid and economical than marking by human graders.
2.1.3.3 Test Components

1) Reading Skills
Both Hughes (1989:116-117) and Heaton (1988:105-106) make a list of reading skills in light of micro and macro skills. Micro skills are exemplified as follows:

-identifying referents of pronouns, etc.
-using context to guess meaning of unfamiliar words
-understanding relations between parts of text by recognizing indicators in discourse, especially for the introduction, development, transition, and conclusion of ideas.

Macro skills are listed as follows, such as:
-Scanning text to locate specific information
-Skimming text to obtain the gist
-Identifying stages of an argument
-Identifying examples presented in support of an argument

According to Weir’s taxonomy of reading, reading skills at the global level include understanding explicitly stated main ideas, inferring prepositional meanings, inferring pragmatic meanings, and skimming for the gist; while reading skills at the local level include inferring lexical meaning, understanding syntax and scanning for specific information (1993: 139).

2) Reading Materials
It is commonly believed that the features of reading materials can determine a reader’s performance in a reading test. The tasks of reading materials should be the basis for the implementation of reading skills. There are two major materials affecting selecting reading materials: one is the nature of the test. Second is objectives of the test. Therefore, the detailed factors which will be taken into consideration in choosing suitable materials can be worked out. They are institutional requirements, teaching and testing specifications, readers’ proficiency levels, the purpose of the readers, reader’s background knowledge, etc., which will all be reflected in the test materials in light of vocabulary, grammatical complexity, organization, topics, cohesion and genres.

3) Reading Speed
As the time limitation in the reading comprehension part is explicitly stated in almost all standard tests, reading speed should be regarded as another component in the reading comprehension test. Nuttal (2000:33) acclaims that “A very slow reader is likely to read with
poor understanding, if only his memory is taxed: the beginning of a paragraph or even a sentence may have been forgotten by the time he has struggled to the end of it” which implies that reading speed and comprehension are closely related.

4) Test formats

Test format is taken to elicit testees’ response. In general, reading techniques are divided into two groups: first, to elicit testees’ selected response. Second, to elicit testees’ constructed response. The typical format for the first group is multiple choice questions (MCQs) while short answer questions (SAQ) and sentence completion (SC) for the second group.

MCQ is deemed to be the most popular item format on account of its convenient implementation and high reliability. However, it does have some innate drawbacks. Weir (1993:44) also argues that “answering multiple-choice items is an unrealistic task as in real life, one is rarely presented with four alternatives from which to make a choice to signal understanding”. In real language context, understanding is acquired through thinking and communicating with each other by speech or writing. Nevertheless, in MCQ there exists guessing. In some circumstances, guessing may result in the reader’s right answer for wrong reasons. Stastically, a reader can get 25% right by guessing in each item. So in some cases, guessing may lead to the reader’s right answer for wrong reasons.

SAQ and SC are the typical representatives of constructed response. Since their format is only allowed for information or data reported in several words or short sentences, they are much more limited in their applicability to wide ranging tests. Several researchers have proved selected response type is easier than the constructed response type in L2 reading comprehension (Bachman 1990:129). The reason is in respect of the requirements on productive skills of the constructed type and fewer chance of guessing than selected response. In this sense, the constructed response type is more representative of real language use. Secondly, in reading comprehension tests, the multiple-choice format is more informative and therefore less appropriate than the open-ended format. Thirdly, the selected response type allows guessing while the constructed response type does not (Hughes 1989:60).
2.2 Theoretical Concerns in Communicative Language Testing

Language and language use are essential for test design and test score interpretation what are embodied in the test. What knowledge of language consists of and how that knowledge is deployed in performance are the foundation of different approaches to language testing. They not only decide what to test or the test construct which are the aspects of knowledge or skill possessed by the candidate which are being measured, but also influence how to test, or the test method adopted.

2.2.1 Bachman’s View of Communicative Language Ability

It is the language ability what testers want to make inferences about from test performance, so it is the central focus of language tests. The definition of the construct which is one of the aspects of language ability to be measured has to base on a valid theory of language ability for any language testing development situation. A theoretical framework of communicative competence is required due to the aim of communicative testing which is the measurement of candidates’ ability to use language for communication.

Bachman extracts knowledge of language from cognitive skills of language use. His framework of communicative language ability is the most comprehensive model up till now (Bachman 1990; Bachman & Palmer 1996). He deems communicative competence to include language competence and strategic competence and language competence consists of “a set of specific knowledge components that are utilized in communication via language” (Bachman 1990:84). Language competence can be divided into two kinds: organizational knowledge and pragmatic knowledge.

Organizational knowledge incorporates knowledge concerning “controlling the formal structure of language for producing or recognizing grammatically correct sentences, comprehending their prepositional content, and ordering them to form text” (Bachman 1990:87). It also comprises of grammatical knowledge and textual knowledge. Meanwhile, grammatical knowledge is the knowledge of vocabulary, morphology, syntax and phonology or graphology for producing or understanding formally accurate utterances, and textual knowledge is the knowledge of conventions for joining utterances together to form a spoken
or written text, which are structured according to rules of cohesion and rhetorical organization. (Bachman 1990:53)

Pragmatic knowledge concerns knowledge “relating utterances or sentences and text to their meanings, to the intentions of language users, and to relevant characteristics of the language use setting” (Bachman & Palmer 1996:69). Two areas of knowledge are also included: illocutionary or functional knowledge which is about “relationships between utterance, texts and the intentions of language users” (Bachman & Palmer 1996:71), and sociolinguistic knowledge of the conventions which determines the appropriateness to a particular language use setting of the conventions. And sociolinguistic knowledge of the conventions determine the appropriateness to a particular language use setting of “the use of dialects or varieties, registers, natural or idiomatic expressions, culture references, and figures or speech” (Bachman & Palmer 1996:53).

2.2.2 Correspondence Between Test Performance and Language Use

The purpose of a communicative language test is to “make inferences about test takers’ ability to use language in a target language use domain” (Bachman & Palmer 1996:44). Test performance must be accordance in demonstrable ways with its target situations (Bachman & Palmer 1996:10). A framework is required to describe the critical features of both language test and non-test language use for this correspondence. And the framework should be constructed on the basis of an examination of the nature of language use.

Language use is “the dynamic and interactive negotiation of intended meanings between two or more individuals in a particular situation” (Bachman & Palmer 1996:61-62). Language users create discourse in using language to express and interpret meanings. The meaning of the discourse is derived from language users’ language ability not only from the discourse itself but more importantly from the relationship between the discourse and the characteristic of the situation and language users.

Consequently, in the construction of a framework of language use for the achievement of the correspondence between test performance and non-test language use, there are two questions to be addressed. One is related to the characteristics of the situation and discourse. The second pertains to the characteristics of the language user, consisting of language ability, topical
knowledge and affective schemata. The first set of characteristics was named by Bachman and Palmer (1996) as task characteristics since they viewed language use as the performance of a set of interrelated language use tasks and believed that language tests should also include language use tasks to ensure their validity.

The two sets of characteristics influence language test performance and language use. The characteristics of individuals, particularly their language ability are related with fidelity and the validity of the inferences made about language ability. The characteristics of the tasks are concerned authenticity, and the validity of the generalization of these inferences to the target domain. Hence, to demonstrate the correspondence between test performance and language use, test developers are required to show the correspondences between both the characteristics of test takers and language users and those of the task and the language use task.

2.2.3 Framework of Task Characteristics

The psychologist J.B. Carroll defines task as “any activity in which a person engages, given an appropriate setting, in order to achieve a specifiable class objectives” (Bachman & Palmer 1996: 43). According to Bachman and Palmer, a language use task is “an activity that involves in using language for the purpose of achieving a particular goal…in a particular situation” (Bachman & Palmer 1996:44). It has three characteristics: a) it is situated in specific situation, b) is goal-oriented c) and involves the active participation of language users.

As Bachman and Palmer (1996: 43-44) argue that a framework of task characteristics is required for testers to select or design test tasks that correspond in specific ways to language use tasks. The goal of communicative tests is to make inferences about test taker’s language ability that generalize to the specific domains in which the test takers are likely to use the language. This domain is called “target language use (TLU) domain” which is a set of specific language use tasks that the test taker is likely to encounter outside the test(44). Because language use, by its very nature, takes place in particular situations, each of which may vary in numerous ways, “each instance of language use is virtually unique”(Bachman & Palmer 1996: 44). Therefore, it is impossible to list all the possible tasks in the description of the target use domain, but it is possible to identify the distinct characteristics of them and to employ them to describe a language use domain. With a framework of task characteristics, target language use tasks and test tasks can be both described and compared with each other.
to assess their degree of correspondence. Thus, the objective to design test tasks that generalize across a range of language use contexts can be attained by “adjusting the characteristics of these test tasks in such a way that they retain the essential features of language use contexts but in focused and idealized form” (Bachman 1990: 111-112). Bachman and Palmer (1996: 47-57) provide a detailed framework of language task characteristics, which can be used in the discourse analysis of language use. It describes five aspects of task: (1) characteristics of setting; (2) characteristics of test rubrics; (3) characteristics of input; (4) characteristics of expected response; (5) relationship between input and response (Bachman & Palmer 1996:47).

(1) Characteristics of the setting
Setting encompasses physical circumstances under which either language use or testing take place. Physical setting, participants and time of task are included.

(2) Characteristics of test rubric
The characteristics of test rubric comprise of the structure of test (number, salience sequence, relative importance of parts), instructions, time allotment and scoring method (e.g. criteria for correctness, procedures for scoring the response, and explicitness of criteria and procedure).

(3) Characteristics of input
Input includes the material contained in a task which the individuals are expected to process and respond to. The format of input has to do with the way in which the input is presented, consisting of channel (aural, visual or both), form (language, non-language, or both), length (single words, phrases, sentences, paragraphs, or extended discourse), type, degree of speededness and vehicle of delivery (live, reproduced via audio or videotape, or both). Characteristics of input contain language characteristics and topical characteristics. Language characteristics includes pragmatic characteristics and organizational characteristics which can be divided into grammatical characteristics and textual characteristics while grammatical characteristics refer to vocabulary, morphology, syntax, phonology, graphology and textual characteristics refer to cohesion, rhetorical or conversational organization. Topical characteristics refer to the type of information such as personal, cultural, academic, or technical that is in the input.

(4) Characteristics of expected response
Expected response is the language use or physical response the test tries to elicit by the instruction, the task, and the input provided. Similar to the input format, the format of expected response can also be described in channel, form, language, length, type and degree of speededness. The only different thing is the type of response: selected response, limited production response and extended production response. The selected response demands test taker to select one response from among two or more provided (multiple-choice being a typical examination example). The limited production response includes a single word or phrase, or even a single sentence or utterance. The extended production response requires the reader to produce an answer longer than a single sentence or utterance. The latter two are viewed as one type by Bachman (1990: 129), the “the constructed type”. The language of expected response is described in terms of the same features as those for the language of the input. Both of the two broad response types have their merits and demerits. The selected type exceeds the constructed type in the following aspects: first, the most significant advantage of the selected response type is its high marking reliability (Hughes 1989: 59). As agreement has already been reached on the correct answer, the marking process is entirely objective and reliable. In addition, selected responses can be scored mechanically by a machine, so the marking is much simpler, rapid and economical than constructed responses, which must be carried out by qualified personnel. Second, since candidates have only to respond by marking on the paper in test requiring selected responses, it is possible to include more items than in those requiring constructed ones (Hughes 1989:60). This means a greater coverage of target reading tasks and a wider sampling of different reading skills, both of which improves the reliability and validity of the test. Third, as now a whole series of statistical tools are available for item analysis, test analysis, score equating, item banking of selected response items, tests of this type can be pre-tested fairly easily (Weir 1993: 43). Although the selected response type is superior to the constructed type in some way, it also has some intrinsic defects and is less useful than the latter. Firstly, as Weir(1993:44) argues, “answering multiple-choice items is an unrealistic task, as in real life one is rarely presented with four alternatives from which to make a choice to signal understanding”. As to the psychological processes involved, the format forces testees to concentrate through convergent thinking on the options provided, which otherwise might not have been thought of. On the contrary, in real language use, understanding of what has been said is achieved through divergent thinking and is communicated through speech or writing if required. In this sense, the constructed response type is more representative of real language use. Secondly, in reading comprehension tests, the multiple-choice format is more informative and therefore less appropriate than the open-ended
format. This is because, according to their investigation of the interrelationship between meaning construction and testing activities, they find that testing activities become information sources which affect test takers’ continuing construction of discourse representation. Thirdly, the selected response type allows guessing, while the constructed response type does not (Weir 1993: 44; Hughes 1989: 60). Statistically, a reader can get 25% right by guessing in each item. So, that is to say, in some cases, guessing may lead to the readers’ right answer for wrong reasons. Fourthly, the selected response type places an unnatural burden on the testees, in that they have to keep in mind four or five potions for each item while they are processing the input (Hughes 1989: 137). Fifthly, item setting is more difficult for multiple-choice tests than for constructed–response ones (Hughes 1989: 61; Weir 1993: 43). Far more items than actually needed have to be written carefully by professional item writers. And it is only after pre-testing and statistical analysis of the performance on the items that the formally usable ones are recognized. Consequently, it is extremely time-consuming and demanding to devise suitable distracters and get the requisite number of satisfactory items. Finally, several researchers in L2 reading comprehension have proved that the multiple-choice response type is easier than the constructed response type (Bachman 1990: 129). This might be explained by some of the features discussed above, such as the requirement of the constructed type on productive skills, and the more extra information and chance of guessing provided by the selected type.

(5) Relationship between input and response

The relationship between input and response is described in terms of the reactivity, scope and directness of the relationship. Reactivity refers to the extent to which the input or the response directly affects subsequent input and response. There are three kinds of tasks of reactivity, namely reciprocal tasks, non-reciprocal tasks and adaptive tasks. When a task engages the individual in interacting with another interlocutor so that the language users can receive feedback from response, it is reciprocal. In a non-reciprocal task, there’s neither feedback nor interaction between language users. In an adaptive task, the performance of the test takers is determined by his or her response to previous tasks. The scope consisting of broad scope and narrow scope demands language users to process tasks based on wider or narrow range of input. A main idea or gist reading comprehension question which deals with the content of an entire passage is an typical examinationple of broad scope test task while the examinationple of narrow broad scope test task is a shot stand-alone multiple choice grammar item that deals with a limited amount of input. Directness of relationship comprising of direct relationship
and indirect relationship refers to the degree to which the expected response can be based mainly on information in the input or whether the language users must depend on information in his own topical knowledge or in the context.

### 2.3 Previous Research

#### 2.3.1 Previous Research on Validity

Validity is one of the most fundamental factors in the large-scale standardized tests. The focus of language assessment and testing research has been on validity since 1960s.

Bachman (1990) believed that validity was the most important quality in the development, interpretation and use of language tests. The framework of task characteristics could be applied to the validation of language testing (Bachman 1990: 152). Bachman studied content and task analysis of the Test of English as a Foreign Language as a first step in investigating the validity of tests. Grotjahn (1986) explored logical task analysis as a basis for making predictions and testing hypotheses related to the study of validity. Perkins and Linnville (1987) investigated construct definition as an approach to construct validation.

Yang and Weir (1998) organized a study on the validity of the National College English Test (CET) in China and achieved with the following results. Firstly, the reliability of CET was high. The reliability coefficients of the objectives items in each test paper were all above 0.9. Secondly, the validity of CET was high. 92% of the teachers thought that CET scores reflect students’ authentic English language ability. Thirdly, the passing scores collected by CET Testing Committee was accordance with what teachers judged. Their correlation coefficient reached 0.82. What’s more, the correlation coefficient was 0.7 between students actual performance in CET and the teachers’ judgment, which was quite rare in large-scale standardized tests. Fourthly, the content design of CET was rational. More than 86% teachers think the content in each test paper was equal for most students from different majors and the allotment of content in each part was appropriate.
2.3.2 Previous Research on TEM-8

Since TEM is a test conducted in China specifically for the Chinese students, the empirical studies of TEM-8 are mainly from China with one exception, i.e. a document from foreign expert which discussed the construct and specifications of TEM-4 and TEM-8 (Ashmore, 1996). Studies of TEM-8 have done by teachers and experts as well as individuals on different parts of TEM-8.

The listening comprehension is the first part of the test. Chen (2003) analyzed the changes in listening of TEM-8 papers from 1996 to 2001. In terms of theories of cognitive psychology and foreign language teaching psychology, he explored the teaching models and methods from the perspective of listening methods and skills in answering question.

The second part of the test is the reading comprehension. Zou, Zhang and Zhou (2002) made an attempt to further validate the potential relationship among question types, test takers’ reading strategies and their reading test scores of TEM-4 reading tests. They discovered the interpretation questions are easiest to elicit students’ reading strategies which were most employed to solve questions at moderate difficult level. Duan and Kong (2004) investigated the reading comprehension in TEM-8 from the ponit of view of textual analysis. They came to the conclusion that when testees encountered problems during reading comprehension process, tactics was suggested so that students could have a better understanding of the passages and increase their scores in TEM-8.

The third part of the test, general knowledge, is a new section dating from 2005, Focusing on reliability, validity and its beneficial or negative backwash effect on the teaching and learning, Li (2007) discussed this part through reviewing and analyzing the examinationination papers from 2005 to 2006.

Proofreading and error correction is on the forth part. Liu (2005) investigates this part from the point of view of cohesion and coherence. According to her, with the help of knowledge on cohesion, and coherence, a lot of errors in this part can be corrected.

The fifth part is translation. Yan and Liu (2002) analyze the errors in students’ translations of TEM-8 examinationination paper of 2002. They believe that to improve students’ translation
skills and abilities, it is the most crucial both in teaching and learning translation theories and underlying their translating practice, basic language skills and raising their awareness of the differences between the English and Chinese languages.

Writing as the last part of TEM-8 test, Liu and Jiang (1998) explore two problems of English writing reflected in TEM-8. One is that English writing has low correlation with other language skills while the other is that English majors studying at school achieved less than those who have graduated. Finally, they provide possible reasons and teaching and writing advices for the two problems.

2.3.3 Previous Research on Reading Comprehension Test (RCT)

Besides validity, there are other variables impacting reading and factors affecting the difficulty of a reading test. In this part, these variables will also be discussed.

Alderson (2000) argues that since there are some variables affecting the nature of reading, they can be divided into two parts. One concerning reader variables are reader’s knowledge, motivation, purpose and other stable reader characteristics. The second one is the test to be read which is called as text variables. Alderson (2000: 60) brings forward the text variables involving text topic and content, text type and genre, traditional linguistic variables, text readability and the medium of the text presentation, etc.

Based on a quantitative analysis of materials and results in the test, Lu (2002) analyzes the variables of RCT. She presents a discussion on the relation between readability, subject matter, item type and the result of a reading comprehension test. Given the significantly discriminating subject matters and different types of test items including dominating factors that lead to the deviation of mean scores, she concludes that a test of reading comprehension with different kinds of subject matters and types of test items can discriminate test takers at different levels.

Wang (2001) conducts a study to explore the similarity and difference in the developing trend and in construction of a diachronic and synchronic comparison and analysis of RCT between CET-4, CET-6, TOFEL and GRE in three aspects: text, stem and option. According to her study, having longer sentences to achieve certain difficulty level which are consistent with
improvement of candidates’ ability is the developing trend of text in RCT. The developing trend of stem in TOEFL’s RCT is to have more stems. Meanwhile, the development of GRE and TOEFL in option length is to have longer stem and short option in accordance with the rule of design option. According to his study, with improvement of the candidates’ comprehension skills, the difficulty level of the test should be increased accordingly.

3 Analysis and Discussion

The analysis and discussion are presented from two perspectives. First, I explore the content validity of the Reading Comprehension Component of TEM-8 from 2008 to 2010 in the terms of the characteristics of input and check whether they fulfill the requirements of *The 2000 Teaching Syllabus for TEM-8* and *The 2004 Test Syllabus for TEM-8*. The characteristics of input comprise of total length, difficulty variables, topics and genre. Reading difficulty is affected by average sentence length, distribution of new information, readability and facility value, and syntactic features. Second, I analyze the content validity of Reading Comprehension of TEM-8 from 2008 to 2010 in the light of expected response based on *The 2000 Teaching Syllabus for TEM-8* and *The 2004 Test Syllabus for TEM-8*. The characteristic of expected response in this thesis refers to the coverage of reading skills embodied in the reading comprehension of TEM-8 examination papers from 2008 to 2010.

3.1 Input

In order to estimate language abilities of a test taker from the score, the materials in the test should meet some requirements. In this study, the passages of reading comprehension of TEM-8 will be analyzed to see to what extent they meet the requirements in the teaching and testing syllabi. Therefore, this part will discuss the characteristics of input of Reading Comprehension comprising of total length, difficulty variables, topics and genres.

3.1.1 Total Length

<table>
<thead>
<tr>
<th>Year</th>
<th>Passage A/Questions</th>
<th>Passage B/Questions</th>
<th>Passage C/Questions</th>
<th>Passage D/Questions</th>
<th>Total words for passages/Questions</th>
<th>Total words</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>634/196</td>
<td>592/214</td>
<td>715/228</td>
<td>584/239</td>
<td>2525/877</td>
<td>3402</td>
</tr>
<tr>
<td>2009</td>
<td>688/215</td>
<td>836/310</td>
<td>524/224</td>
<td>713/183</td>
<td>2761/932</td>
<td>3693</td>
</tr>
<tr>
<td>2010</td>
<td>1401/335</td>
<td>698/175</td>
<td>471/300</td>
<td>600/151</td>
<td>3170/961</td>
<td>4131</td>
</tr>
</tbody>
</table>
As demonstrated in the table, the word number of each passage differs slightly. The most noticeable feature is that the number of total words for passages is increasing year after year, ranging from 2525 to 3170. At the same time, the number of total words for questions ranges from 877 to 961. It is necessary to evaluate whether the numbers in Table 3 are consistent with the specific requirements of syllabi which consist of speed limitation.

_The 2000 Teaching Syllabus:_ To be able to read at the speed of 140-180 words per minute (wpm.) (Higher Education Institution, 2000: 7).

_The 2004 Test Syllabus:_ The total length of passages and the following questions should be about 3000 words, which are to be processed and responded to within 30 minutes (Higher Education Revision Group, 2004: 6).

Being able to read at a speed of 140-180 wpm refers to the ability of reading passages in relatively relaxed circumstances without answering questions of the passages. For instance, when students read some novels or newspapers, their reading speed is required to 140-180 wpm. So the total words of their reading should be 4200 to 5400 words within 30 minutes. Nevertheless, this requirement is not suitable in TEM-8. First, the test takers are under considerable high pressure and they must read more carefully and try to make sense of each word since there are several MCQs followed. Second, some difficult sentences or paragraphs should be reread. When doing MCQs of related passages, in order to ensure the accuracy, students often go back to search the passages for particular information. Finally, blackening the answer sheet also requires time. As MCQs are written on an answer sheet to be read mechanically, blackening the items on the answering sheet also needs time to do. Thus the test syllabus for TEM-8 requires 3000 words in total, which has already taken into consideration of the time spent on thinking over the questions, rereading some of paragraphs and blackening the answers. Therefore, the figures of 2010 in the table above do not correspond to the requirements of the teaching syllabus or the test syllabus.

### 3.1.2 Difficulty Variables

Test takers must be able to comprehend book reviews and editorials in newspapers and magazines in British or American English, as well as literary works and biographies to a certain degree of difficulty. But syllabi do not include specific information about the degree of difficulty and what should contribute the degree of difficulty. So variables consisting of
average sentence length, distribution of new information, readability and facility value and sentence types are used to investigate the degree of difficulty and their consistency to each examination paper of Reading Comprehension of TEM-8 from 2008 to 2010.

3.1.2.1 Average Sentence Length

Despite the absence of specific requirement as regards in the syllabi, it is necessary to study whether there is a discrepancy of sentence length difficulty in the three papers as long sentences can block comprehension even when the vocabulary is familiar (Nuttall 2000:78). The calculation of the average sentence length of each passage is listed in the following table.

Table 4: The average sentence length of reading passages from 2008 to 2010

<table>
<thead>
<tr>
<th>Year</th>
<th>Total words</th>
<th>Total sentences</th>
<th>Average sentence length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008(A)</td>
<td>634</td>
<td>31</td>
<td>20.45</td>
</tr>
<tr>
<td>2008(B)</td>
<td>592</td>
<td>31</td>
<td>19.10</td>
</tr>
<tr>
<td>2008(C)</td>
<td>715</td>
<td>38</td>
<td>18.82</td>
</tr>
<tr>
<td>2008(D)</td>
<td>584</td>
<td>43</td>
<td>13.58</td>
</tr>
<tr>
<td>Total</td>
<td>2525</td>
<td>143</td>
<td>17.66</td>
</tr>
<tr>
<td>2009(A)</td>
<td>688</td>
<td>27</td>
<td>24.48</td>
</tr>
<tr>
<td>2009(B)</td>
<td>836</td>
<td>41</td>
<td>20.39</td>
</tr>
<tr>
<td>2009(C)</td>
<td>524</td>
<td>20</td>
<td>26.20</td>
</tr>
<tr>
<td>2009(D)</td>
<td>713</td>
<td>25</td>
<td>28.52</td>
</tr>
<tr>
<td>Total</td>
<td>2761</td>
<td>113</td>
<td>24.43</td>
</tr>
<tr>
<td>2010(A)</td>
<td>1401</td>
<td>53</td>
<td>26.43</td>
</tr>
<tr>
<td>2010(B)</td>
<td>698</td>
<td>30</td>
<td>23.27</td>
</tr>
<tr>
<td>2010(C)</td>
<td>471</td>
<td>15</td>
<td>31.4</td>
</tr>
<tr>
<td>2010(D)</td>
<td>600</td>
<td>19</td>
<td>31.58</td>
</tr>
<tr>
<td>Total</td>
<td>3170</td>
<td>117</td>
<td>27.09</td>
</tr>
</tbody>
</table>

As it is shown in the above table, the average sentence length in the three years ranges from 2008 to 2010 ranges from 17.66 to 27.09. The longest average sentence length is the paper from 2010, which indicates that there are more complex sentence types. So it seems the passages in 2010 are more difficult to go through within the same time limit when considering average sentence length.
The difference of average sentence length of the four passages in each paper should be in a relative balance so testees will not feel abrupt to read from one passage to another one. As the above table demonstrates, it can be found that the average sentence length of the four passages in 2008 varies least, from 13.58 to 20.45. Nevertheless, the difference in the other two average sentence length in each passage in 2010 is more prominent compared with that in other two papers. Since passage A in 2010 is a literary story, the average sentence length is relatively longer.

Generally speaking, the layout of reading comprehension part concerning average sentence length is comparatively consistent in the three years, which can guarantee a certain degree of difficulty level.

3.1.2.2 Distribution of New Information

There are two types of distribution of new information in the passages, namely, compact and diffuse. If the passage contains many a new information concerning its length, the passage is considered comparatively compact while holding relatively a little new information to its length is diffuse. To shed some light on distribution of new information, calculating the ratio of content words comprising of nouns, full verbs, adjectives and adverbs to the total words is used here (Bachman et al. 1988). Thus, counting the content words is the first step to check the validity. The table of content words in each passage is listed in the following.

Table 5: Content words in reading passages from 2008 to 2010

<table>
<thead>
<tr>
<th>Passages</th>
<th>Content words</th>
<th>Total words</th>
<th>Percentage of content words</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008(A)</td>
<td>337</td>
<td>634</td>
<td>53.15%</td>
</tr>
<tr>
<td>2008(B)</td>
<td>343</td>
<td>592</td>
<td>57.94%</td>
</tr>
<tr>
<td>2008(C)</td>
<td>484</td>
<td>715</td>
<td>67.69%</td>
</tr>
<tr>
<td>2008(D)</td>
<td>309</td>
<td>584</td>
<td>52.91%</td>
</tr>
<tr>
<td>Total</td>
<td>1473</td>
<td>2525</td>
<td>58.34%</td>
</tr>
<tr>
<td>2009(A)</td>
<td>371</td>
<td>688</td>
<td>53.92%</td>
</tr>
<tr>
<td>2009(B)</td>
<td>447</td>
<td>836</td>
<td>53.47%</td>
</tr>
<tr>
<td>2009(C)</td>
<td>290</td>
<td>524</td>
<td>55.34%</td>
</tr>
<tr>
<td>2009(D)</td>
<td>374</td>
<td>713</td>
<td>52.45%</td>
</tr>
<tr>
<td>Total</td>
<td>1482</td>
<td>2761</td>
<td>53.68%</td>
</tr>
<tr>
<td>2010(A)</td>
<td>805</td>
<td>1401</td>
<td>57.46%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>2010(B)</td>
<td>435</td>
<td>698</td>
<td>62.32%</td>
</tr>
<tr>
<td>2010(C)</td>
<td>257</td>
<td>471</td>
<td>54.56%</td>
</tr>
<tr>
<td>2010(D)</td>
<td>341</td>
<td>600</td>
<td>56.83%</td>
</tr>
<tr>
<td>Total</td>
<td>1838</td>
<td>3170</td>
<td>57.98%</td>
</tr>
</tbody>
</table>

Take a thorough look at the table above, it can be found that the three papers are pretty consistent with regard to the distribution of new information. The percentage of content words ranges from 50% to 60%, which does not vary much from each year and illustrates a relative proper intensity of new information according to the percentage. Therefore, the above table depicts a rather stable picture for the distribution of new information.

Bachman (1990:135) believes that since highly compact or highly diffuse distribution of information is too demanding for test takers’ competence, this kind of input of highly compact or highly diffuse distribution of information is not expected in reading passages in order to correspond to the syllabi. Highly compact input will not allow much opportunity for test takers to process meaning efficiently when highly diffuse information distribution requires test takers to encode information or process some information by themselves. So input of neither too compact or too diffuse is more acceptable in the achievement test, which provides more opportunity for test takers to negotiate meanings. Thus, the distribution of new information in the three papers corresponds to syllabi of TEM-8. As a variable influences the difficulty of reading comprehension, the consistency of the new information distribution in each paper has guaranteed the moderate validity.

### 3.1.2.3 Readability

Readability means “how easily written materials can be read and understood”, which “depends on many factors, including (a) the average length of sentences in a passage; (b) the number of new words a passage contains; (c) the grammatical complexity of the language used. Procedures used for measuring readability are known as ‘readability formulae’ ” (Richards & Schmidt. 2003: 442).

There are many kinds of readability formulas such as Flesch readability formula, Lorge readability formula and Bormuth readability formula, in which the Flesch Formula is the most widely known and frequently used formula both in the world. The formula is $RE=206.84 -$
Here $wl$ represents the average number of syllables in 100 words, $sl$ refers to the average number of words per sentence. The scales of readability scores are listed in Table 7. The higher the readability score, the easier the passage. It is also worth noting that the readability scores may not show the actual difficulty of these passages but are regarded them as reference proof. In Microsoft Word, the formula for calculating the Flesch Reading Ease score is $206.835 - (1.015 \times ASL) - (84.6 \times ASW)$ (Microsoft office 2003), which has the same function with the most widely used formula mentioned above. ASL represents the average number of words per sentence while ASW refers to the number of syllables per word. Thus, the readability scores in this thesis can be calculated by the tool in Microsoft Word. The readability of each passage in these three papers is shown in the following table.

**Table 6: The readability of reading passages from 2008 to 2010**

<table>
<thead>
<tr>
<th>Passage Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Average readability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>49.9</td>
<td>51.5</td>
<td>51.7</td>
<td>66.1</td>
<td>54.8</td>
</tr>
<tr>
<td>2009</td>
<td>67.6</td>
<td>63.5</td>
<td>62.4</td>
<td>72.2</td>
<td>66.6</td>
</tr>
<tr>
<td>2010</td>
<td>48.7</td>
<td>49.2</td>
<td>51.5</td>
<td>43.3</td>
<td>48.2</td>
</tr>
</tbody>
</table>

**Table 7: Reference table of readability (Flesch, 1948: 221)**

<table>
<thead>
<tr>
<th>Readability score</th>
<th>Description of style</th>
<th>Typical magazine</th>
<th>Potential readers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-30</td>
<td>Very difficult</td>
<td>Scientific</td>
<td>College</td>
</tr>
<tr>
<td>30-50</td>
<td>Difficult</td>
<td>Academic</td>
<td>High school</td>
</tr>
<tr>
<td>50-60</td>
<td>Fairly difficult</td>
<td>Quality</td>
<td>Some H.S.</td>
</tr>
<tr>
<td>60-70</td>
<td>Standard</td>
<td>Digest</td>
<td>7th or 8th grade</td>
</tr>
<tr>
<td>70-80</td>
<td>Fairly easy</td>
<td>Slick-fiction</td>
<td>6th</td>
</tr>
<tr>
<td>80-90</td>
<td>Easy</td>
<td>Pulp-fiction</td>
<td>5th</td>
</tr>
<tr>
<td>90-100</td>
<td>Very easy</td>
<td>Comics</td>
<td>4th</td>
</tr>
</tbody>
</table>

As shown in the Table 6 and compared to Table 7 above, the passages from 2010 are the most difficult to comprehend while those in 2009 are the easiest. With a reference to the Flesch readability standard, it is also can be found that most of the passages are within the scales from
30 to 50 and 50 to 60 which refers to “difficult” and “fairly difficult” respectively except one passage which exceeds the desirable scale of readability. It also indicates that the testees at the undergraduate level have the similar difficult level of native students at between 7th or 8th and high school. Meanwhile, it also shows that the difficulty level of Reading Comprehension in the three papers is set properly in general which accords to the requirement of the syllabi.

### 3.1.2.4 Sentence Types

Due to the time and space limitation, the organizational features of Reading Comprehension of TEM-8 from 2008 to 2010 will be explored in terms of sentence type as one grammatical category. In general, sentences can be divided into four types, namely, simple, complex, compound and compound-complex. Quirk et al. (1985) points out that simple sentence consists of only independent clause that does not have another clause as one of its elements, e.g. “At the age of 16, Lee Hyuk Joon’s life is a living hell.”(S1 from passage A in 2008); complex sentence refers to that which only have one main clause but have one or more subordinate clauses as an element of the sentence, for instance, “That’s because South Korea has conceived a new college-entrance system, which will be implemented in 2008”.(S10 from passage A in 2008); compound sentence refers to that which comprises two or more coordinated main clauses, for instance, “It’s a grueling schedule, but Lee worries that it may not be good enough to get him into a top university.”(S5 from passage A in 2008); compound-complex sentence is the sentence with two or more coordinated main clauses and one or two more subordinate clauses, for instance, “Nearly all Korean high school will be weighted equally in the college-entrance process, and relatively weak students in provincial schools, who may not score well on standardized tests, often compile good grade-point averages” (S37 from passage A in 2008). The table 8 shows the number of the four types of sentences and their percentage in each passage.

<table>
<thead>
<tr>
<th>Passages</th>
<th>Simple</th>
<th>Complex</th>
<th>Compound</th>
<th>Compound-complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008(A)</td>
<td>16/51.61%</td>
<td>11/35.48%</td>
<td>3/9.68%</td>
<td>1/3.23%</td>
</tr>
<tr>
<td>2008(B)</td>
<td>11/35.48%</td>
<td>16/51.61%</td>
<td>2/6.45%</td>
<td>2/6.45%</td>
</tr>
<tr>
<td>2008(C)</td>
<td>15/39.47%</td>
<td>11/28.95%</td>
<td>8/21.05%</td>
<td>4/10.53%</td>
</tr>
<tr>
<td>2008(D)</td>
<td>23/53.49%</td>
<td>12/27.91%</td>
<td>2/4.65%</td>
<td>6/13.95%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2009(A)</td>
<td>2009(B)</td>
<td>2009(C)</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>65/57.52%</td>
<td>9/33.33%</td>
<td>19/46.34%</td>
<td>1/5.00%</td>
</tr>
<tr>
<td></td>
<td>50/34.97%</td>
<td>14/51.85%</td>
<td>18/43.90%</td>
<td>17/85.00%</td>
</tr>
<tr>
<td></td>
<td>15/10.49%</td>
<td>3/11.11%</td>
<td>1/2.44%</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td>13/9.09%</td>
<td>2/7.41%</td>
<td>3/7.32%</td>
<td>2/10.00%</td>
</tr>
<tr>
<td></td>
<td>35/30.97%</td>
<td>38/32.48%</td>
<td>14/26.42%</td>
<td>12/40.00%</td>
</tr>
<tr>
<td></td>
<td>60/53.10%</td>
<td>35/66.04%</td>
<td>16/53.33%</td>
<td>5/33.33%</td>
</tr>
<tr>
<td></td>
<td>7/6.19%</td>
<td>2/3.77%</td>
<td>1/1.89%</td>
<td>7/46.67%</td>
</tr>
<tr>
<td></td>
<td>12/10.62%</td>
<td>2/3.77%</td>
<td>1/1.89%</td>
<td>2/3.77%</td>
</tr>
<tr>
<td></td>
<td>38/32.48%</td>
<td>7/36.84%</td>
<td>10/52.63%</td>
<td>5/33.33%</td>
</tr>
<tr>
<td></td>
<td>68/58.12%</td>
<td>10/52.63%</td>
<td>15/10.49%</td>
<td>7/46.67%</td>
</tr>
<tr>
<td></td>
<td>6/5.13%</td>
<td>1/5.26%</td>
<td>13/9.09%</td>
<td>2/7.41%</td>
</tr>
<tr>
<td></td>
<td>5/4.27%</td>
<td>13/9.09%</td>
<td>2/7.41%</td>
<td>5/4.27%</td>
</tr>
</tbody>
</table>

From the table above, it demonstrates that the same sentence type bears a quite similar proportion in the passages of 2009 and 2010 except the proportion of complex-compound sentences in 2009 which is higher than that in 2010, from 10.62% to 4.27%. The simple sentences range from 30.97% to 57.52% in three years and the complex sentences range from 34.97% to 58.12%, which suggests that the two types are the main sentences types in the TEM-8 Reading Comprehension. Compound sentences which are the combination of two simple sentences and are not very difficult to understand, occupy the highest percentage, 10.49% in 2008 compared to the 6.19% in 2009 and 10.62% in 2010 while the percentage of compound-complex sentences which are the most complex sentences of the four types, are pretty close with the number in 2009, ranging from 9.09% to 10.62%. Besides, since the simple sentences claim 57.52% in 2008 compared to the number of 30.97% and 32.48% respectively in 2009 and 2010 while the complex sentences claim 34.97% in 2008 compared to the number of 53.10% and 58.12% in 2009 and 2010 respectively, the sentences in passages of 2008 are simpler than those of 2009 and 2010. In summary, it has also been established that among these three papers, the Reading Comprehension in 2009 occupies the most complex sentences in all. Take a general look at the figures in the table above, we can find the percentage of complex sentence types are increasing year by year, which suggests the difficulty level is also increasing slightly.
3.1.3 Topics

According to Richards et al. (2003: 557), topic refers to “what is talked about or written about”. Weir (1993: 67) believes that “the topic should be selected from a suitable genre at an appropriate level of specificity, and should not be culturally biased or favor one section of the test population”. The 2004 test syllabus also emphasizes that range of topics should be wide, including society, science and technology, culture, economy, daily life, biography and so on. In this study, topics will be classified into eight groups according to the examples given in the syllabi, namely, education, social affairs, cultural affairs, literary works, personal experience.

<table>
<thead>
<tr>
<th>Passage</th>
<th>Topic</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008(A)</td>
<td>The education system in South Korea</td>
<td>Education</td>
</tr>
<tr>
<td>2008(B)</td>
<td>Migration back to farmland in UK</td>
<td>Social affairs</td>
</tr>
<tr>
<td>2008(C)</td>
<td>Castellers in Barcelona</td>
<td>Cultural affairs</td>
</tr>
<tr>
<td>2008(D)</td>
<td>A story of the bankruptcy of a law firm</td>
<td>Literary works</td>
</tr>
<tr>
<td>2009(A)</td>
<td>Travelling to Istanbul with children</td>
<td>Personal experience</td>
</tr>
<tr>
<td>2009(B)</td>
<td>Get the retired workforce back to company</td>
<td>Social affairs</td>
</tr>
<tr>
<td>2009(C)</td>
<td>The problems of working wives</td>
<td>Social affairs</td>
</tr>
<tr>
<td>2009(D)</td>
<td>The pilgrimage to Tibetan</td>
<td>Cultural affairs</td>
</tr>
<tr>
<td>2010(A)</td>
<td>The rickshaw in Calcutta</td>
<td>Cultural affairs</td>
</tr>
<tr>
<td>2010(B)</td>
<td>Waiting in lines in America</td>
<td>Social affairs</td>
</tr>
<tr>
<td>2010(C)</td>
<td>The luxury in Bbylonian hotel</td>
<td>Literary works</td>
</tr>
<tr>
<td>2010(D)</td>
<td>The Alcoa project with Iceland</td>
<td>Social affairs</td>
</tr>
</tbody>
</table>

Alderson (2000:61) says that “what readers know will affect what they understand when reading, so it is commonly assumed that text content will affect how readers process text”. From Table 9, the topic of social affairs and cultural affairs are covered the most. Judging from the distribution, we can know topics are centralized on the social affairs and cultural affairs. In each paper, there are social affairs and cultural affairs but the allotment of topics in 2008 distributes better than those in 2009 and 2010 since the passages in 2008 cover four topics while in 2009 and 2010 there are two passages of social affairs respectively.

According to Weir (1993), reading material in a test should be novel and interesting to test each level of reading ability and should contain enough information to raise questions. And the most
important point is that reading material in a test should not be too professional which will give more advantage to some students over others because of familiarity. And Alderson (2000: 102) discusses the importance of selecting proper materials. “Whilst one might claim that background knowledge in the content area of a reading text can enable students to perform to the best of their ability, tests based on texts which are too specialized might test subject matter knowledge rather than reading ability. Subject-related texts might also discriminate against individuals who happen to possess less background knowledge in a particular field”. As it is shown in the table, passages in TEM are not quite professional but of various topics. It is advisable that passages selected in TEM-8 should cover more various topics to better assess students’ English reading ability.

### 3.1.4 Genres

The term “genre” refers to “a type of discourse that occurs in a particular setting, that has distinctive and recognizable patterns and norms of organization and structure, and that has particular and distinctive communicative functions” (Richards et al. 2003:224). According to Ding, et al. (1994), types of genre can be categorized into four types, namely, description, narration, exposition and argumentation. Description means to describe or present a person, place, event or action, which often appears in journal writing and poetry. Narration is employed to tell a story or narrate an event. Exposition is a kind of genre to convey factual information, to clarify reasons, to deliver instructions or to express feelings and opinions. Argumentation is to analyze and comment on some issues or affairs, or try to persuade, convince, to change another’s viewpoint or opinion or thinking. The genres of passages in TEM-8 of 2008 to 2010 are listed in the following table.

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
<th>Narration</th>
<th>Exposition</th>
<th>Argumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>(%)</td>
<td>16.67%</td>
<td>33.33%</td>
<td>0</td>
<td>50%</td>
</tr>
</tbody>
</table>

From the table above, it can be seen that exposition is excluded from the Reading Comprehension of the three years. A possible explanation is that exposition is always have clear
clues and is relatively more effortless to comprehend, which assists readers to achieve information quickly. And there are only two kinds of genres in 2008, which is not very consistent with the requirements of genres set by syllabi. In The 2004 Test Syllabus, it is required that “A variety of text genres such as: narration, description, exposition, argumentation, advertisements, instructions, diagrams, etc.”, which means types of passages should be diversified, not only including the above mentioned types but also advertisements, instructions and diagrams. Compared with the distribution in three years, it is obvious that the argumentation is the first mostly used genre which holds 50% of the total passages. Narration comes the second which takes up 33.33% while description only accounts for 16.67%. Adopting more argumentative passages may be intended to test students’ abilities of processing and analyzing information. The 2000 Teaching Syllabus claims that students of English major should be of the ability to obtain information and utilize the information, the ability to analyze questions and give their own ideas independently and the ability to be creative. More argumentative reading materials may contribute to cultivating the kinds of abilities. Narration is often to test students’ another ability, namely, acquiring information in the passages, which means in the Table 10, the ability of acquiring information is also an important aspect to achieve in the Reading Comprehension of TEM-8. Since The 2004 Testing Syllabus serves as a blueprint for designing test papers, claiming that genres of reading passages should be various, so it is suggested to cover more other genres including expositions, advertisements, charts and diagrams in TEM-8.

3.2 Expected Response

According to Bachman and Palmer (1996: 53), expected response refers to the language use or physical response which the tests attempt to elicit by means of the instructions, the tasks and the input provided. The format of expected response refers to the way in which the response is produced. As established above under the characteristics of input, the characteristics of expected response also include length, propositional content, topics, genres. However, there is only one type of expected response employed in the Reading Comprehension of TEM-8 tests so other characteristics do not need to be discussed. In Reading Comprehension of TEM-8, the MCQs are used to elicit the selected response. Attention has been paid to the coverage of reading skills in this study. According to Alderson (2000), readers not only have knowledge but also have abilities which not only to acquire new knowledge but also to process information. Researchers have long concerned that readers may have acquired the abilities or skills to process text and to
answer particular questions. It is requisite to discuss reading skills in TEM-8. *The 2004 Testing Syllabus* (Higher Education Revision Group, 2004: 17) defines the specific requirement of the reading skills for TEM-8 as follows:

1. Understanding the main ideas by careful reading and distinguishing facts and detailed information
2. Understanding the literal meaning and implicitly stated information
3. Understanding and deducing the implied meaning
4. Understanding the author’s attitude and intention, discourse structures, language characteristics and rhetorical devices

Zou, Zhang & Zhou (2002) believe that reading skills required can be manifested by the question itself. Many have attempted to classify the different question types in reading test (Smith, 1978 & Nuttle 1982 & Alderson, 2000). In this thesis, Nuttle’s category of the question types is taken to check whether the test design of TEM-8 includes all the types and to examine their distribution in the tests, their proportion to each other and the consistency over the three papers. Nuttle (2000: 187-88) categorizes the questions into six types:

**Type 1**
Questions demanding literal comprehension, whose answers are directly and explicitly available in the text.

**Type 2**
Questions involving reorganization or reinterpretation, which require student either to reinterpret literal information or to obtain it from various parts of the text and put in together in a new way, perhaps using elementary inferencing.

**Type 3**
Questions requiring identification of inference, which oblige students to consider what is implied but not explicitly state and require students to put together the pieces of information that are scattered throughout the text.

**Type 4**
Questions demanding evaluations which ask for a considered judgment about the text terms of what the writer is trying to do, and how far he/she has achieve it.
Type 5
Questions requiring personal response, which ask for the readers own reactions or ideas to the text not under the writer’s response.

Type 6
Questions concerned with how writers say what they mean, which is intended to give students strategies for handling the text in general, rather than help them to understand one particular text.

As the MCQs demand receptive skills rather than productive skills and Type 5 concerns with productive skills, Type 5 can thus be excluded from the analysis as testees are not allowed to display their productive skills from the Reading Comprehension of TEM-8. Thus, the reading skills covered in the syllabus for TEM-8 will be explored through question stems in which the reading skills embodied. The first step is to categorize the question stems in terms of specific reading skills illustrated in syllabus into different types.

The reading skills involving comprehending the literal meaning or clearly showed information in the text can be reflected through Type 1. For instance, “In the last but two paragraphs, the Ramblas is described as “a front-row seat for Barcelona’s longest running theatrical event”. What does it mean?” (2008, No. 24) and “‘The company spends a lot of time on the ergonomics of its factories’ (Paragraph Seven) means that _ _ _”.

The reading skills involving distinguishing and reinterpreting detailed information and facts can be reflected through Type 2. For instance, “According to the passage, the new college-entrance system is designed to _ _ _.” (2008, No. 11), “According to critics, the popularity of private education is mainly the result of _ _ _.” (2008, No. 13)

As for Type 3, there are usually some signal words in this type such as “seem, imply, suggest, indicate, infer, etc.”, which oblige students to consider what is implied For instance, “What seems to be the effect of introducing the new system?”(2008, No. 12) and “The sentence in the
last paragraph ‘... Britain isn't catching up with mainland Europe; it's leading the way’ implies that ___ ___” (2008, No. 20)

The reading skills required for understanding the main ideas, authors’ attitudes and intentions can be reflected in Type 4. For examinationple, “Which of the following adjectives best describes the author’s treatment of the topic?” (2008, No. 15) and “What is the main impression of the scenes on the Ramblas?” (2008, No. 25).

Type 5 concerns the skills of understanding discourse structure, language characteristics and rhetorical devices. Such as “Which of the following statements contains a metaphor?”, (2008, No. 27). “Which of the following best describes the author’s development of argument?” (2009, No. 20).

Table 11: Question types identified in selected reading passages from TEM-8 (2008 - 2010)

<table>
<thead>
<tr>
<th>Level&amp;Type</th>
<th>Year</th>
<th>Low-order skills</th>
<th></th>
<th></th>
<th>High-order skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Type 1</td>
<td>Type 2</td>
<td>Type 3</td>
<td>Type 4</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11/55%</td>
<td>9/45%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>11/55%</td>
<td>9/45%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>12/60%</td>
<td>8/40%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>28</td>
<td>14</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>34/56.67%</td>
<td>26/43.33%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These question types are all distributed at different proportion in each paper but without great distribution gap. The question types can be further divided into two types, namely, those demanding the low-order skills and those requiring high-order ones (Yang & Weir 1998: 71). The skills of comprehending the literal meaning or clearly manifested information and distinguishing or reinterpreting the detailed information through context clues belong to the low-order skills while the skills of comprehending implicitly stated information, understanding the topic and main idea, guessing a writer’s intention and attitudes belong to the high-order skills. Low-order skills are thus embodied in Question Type 1 and Type 2 whereas the high-order skills are represented in Question Types 3, 4 and 6. Consequently, the results can be
discussed in two ways: one is discussed from the distribution variance in the three papers; the other is from the overall characteristics of the skill coverage in the three papers.

Table 11 demonstrates that there are 34 items that involve with the reading skills at the low-order level occupying 56.67% of the total while the other 26 items are dealt with the reading skills at the high-order level, which implies that testees who only master the reading skills of low-order level will barely pass the examination. However, if testees attempt to gain high scores of this part, they are required to acquire more advanced skills, such as comprehending the implied meaning of passages. Table 11 shows that Question Type 2 has the highest percentage and Question Type 3 comes the next highest. Being the most basic and fundamental requirement of reading comprehension, Question Type 2 is the representative of the skills at low-order level testing whether testees have understood the detailed information or facts.

Question Type 3 requires more advanced skills since it measures the readers’ ability to understand the implied meaning. Question Type 1, 4 and 6 share the same percentage in the three years in which Question Type 4 is meant to guess writer’s intention which demands thoroughly understanding the whole text so Question Type 4 is more difficult than Question Type 3 in the reading comprehension part of TEM-8. Question 1 is the easiest to tackle since it only requires the literal comprehension. Question Type 6 is the most difficult as it requires the awareness of the author’s writing techniques. Consequently, Question Type 1, 4 and 6 account for a smaller proportion in the three years in order to balance the difficulty level in TEM-8

As for the difference in the distribution of skills in the three papers, it is found that the 2010 paper comes first in terms of question type demanding the low-order skills when the other two papers turn out to be the same. Nevertheless, the readability of 2010 is the lowest among the three papers, which means the it is the most difficult according to the Table 7. Reasons are given that test takers may encounter some complex sentences in reading passages of 2010 but the question items are not so difficult to answer so there is a compromise in it, which makes the test of 2010 not so difficult to deal with. As a result, the different requirements for the reading skills as high-order or low-order on their own can not determine the level of difficulty of the tasks set on the different passages. The reading materials should also be taken into consideration. Therefore, it is justifiable that the features of the reading material can largely determine a reader’s performance in a reading test, which corresponds to Xin’s finding (2007) who finds that if a stem of question is a detailed question but the detailed information for this
question is difficult to find from the text, the question is difficult to answer; if a stem is an inference question, the answer is easy to infer, the question is easy.

4. Summary and Conclusion

In this section, major findings, a brief discussion of implications and some limitations of this study are presented.

4.1 Major Findings

Based on The 2000 Teaching Syllabus for TEM-8 and The 2004 Test Syllabus for TEM-8, the Reading Comprehension of TEM-8 from 2008 to 2010 is a valid reading test by the content analysis.

(a) The characteristics of input of Reading Comprehension of TEM-8 in terms of total length in each paper generally correspond to the requirements of syllabi. But the total word number of passages and questions is inclined to increase within a reasonable range except for the paper in 2010 whose total word number is a little beyond the requirement set by the syllabus. Meanwhile, it also indicates the trend of total words number will be inclined to increase in recent years because of social demands which require people to obtain more and more information in this information age.

(b) The characteristics of TEM-8 and the requirements of syllabi of TEM-8 can be well reflected in the difficulty variables regarding to the characteristics of input of Reading Comprehension of TEM-8. To begin with, the average sentence length of Reading Comprehension in the three years is in a well balance which can guarantee that test-takers will not feel it abrupt to read passages sentence by sentence. What’s more, the three papers have good consistency with regard to the distribution of new information. The percentage of content words to the total words ranges from 50% to 60%, which is neither too compact nor too diffuse. Third, the difficult level of Reading Comprehension corresponds to the requirements of the syllabi. The readability of most passages is within the scale 60 to 30, which is proved to set properly in general. Ultimately, the sentence types in 2009 and 2010 are consistent to each other, which indicates that the difficulty of the syntax is under a reasonable control while the sentence types in 2008 is simpler than the other two. As for the percentage of the sentence of complex type is increasing year by year, it infers the difficulty level in terms of sentence types is increasing year by year.
(c) Being one of the characteristics of input of Reading Comprehension in an equal distribution, topics in the three papers cover various fields. At the same time, more various topics are suggested to bring in the tests of TEM-8 to encourage students to read extensively in the reading training, which is more consistent with the requirement of subject specificity for English majors.

(d) Genre, as one characteristic of input of Reading Comprehension of TEM-8 basically corresponds to the requirements of syllabi. But other genres are encouraged to enrich the genres of passages of TEM-8, for instance, advertisements, instructions and digrams.

(e) The five question types which are the characteristic of expected response, serve as a way to manifest the coverage of reading skills. Though the distribution and percentage of these skills in each paper are different but are under a reasonable control.

4.2 Implications

According to the findings obtained in the theoretical and practical aspects of the research, three implications are provided for the improvement of test design.

First, though MCQ is usually regarded as the common question type to ensure the objectivity and impartiality of the scores and the stability of the test which guarantee the reliability of the test, desirable correspondence between test performance and authentic language use can not be gained. Therefore, it is suggested to enrich more test formats in the reading assessments. Zou (1997) proposes constructed response to be employed in Reading Comprehension of TEM-8 in the future as an attempt to fulfill the requirements of The 2000 Teaching Syllabus because of some unavoidable defects of selected response type as I suggested in characteristics of expected response in 2.2.3. A more use-oriented and comprehensive model of language competence is encouraged to adopted in TEM-8 in the future.

Second, the total words of in the passages in Reading Comprehension also should be careful. Beyond some limitation of total words, students will not be able to complete the whole reading comprehension task or not be able to finish it successfully as students not only have to finish the reading materials but also have to answer the following questions in a limited time.
Third, as discussed in the major findings, there is a lack of various genres in the Reading Comprehension of TEM-8, it is advisable that a proper increase of the exposition and other genres such as advertisements, instructions and charts are encouraged to bring in.

Fourth,

4.3 Limitations

Firstly, this thesis focuses solely on exploring the content validity of Reading Comprehension from 2008 to 2010 in terms of characteristics of input and expected response. And only content validity of Reading Comprehension of TEM-8 is explored in this thesis by reason of time and energy limitation and sufficient data available.

Second, it results in the impossibility of calculating F.V. of each question because of lack of testee’s performance scores.

Third, this thesis combining the nature of reading, builds a new framework on the basis of the framework of the task characteristics proposed by Bachman and Palmer (1996). But the new one also has some deficiencies such as, some characteristics of input like cohesion devices, figurative speeches and key words in the passages for full comprehension have not analyzed because they offer a lot of difficulty to measure.
References

Secondary Sources


Appendices

Appendix 1

TEST FOR ENGLISH MAJORS (2008)
-GRADE EIGHT-

PART II READING COMPREHENSION (30 MIN)

In this section there are four reading passages followed by a total of 20 multiple-choice questions. Read the passages and then mark your answers on your coloured answer sheet.

TEXT A

At the age of 16, Lee Hyuk Joon's life is a living hell. The South Korean 10th grader gets up at 6 in the morning to go to school, and studies most of the day until returning home at 6 p. m. After dinner, it's time to hit the books again – at one of Seoul's many so-called cram schools. Lee gets back home at 1 in the morning, sleeps less than five hours, then repeats the routine – five days a week. It's a grueling schedule, but Lee worries that it may not be good enough to get him into a top university. Some of his classmates study even harder.

South Korea's education system has long been highly competitive. But for Lee and the other 700,000 high-school sophomores in the country, high-school studies have gotten even more intense. That's because South Korea has conceived a new college-entrance system, which will be implemented in 2008. This year's 10th graders will be the first group evaluated by the new admissions standard, which places more emphasis on grades in the three years of high school and less on nationwide SAT-style and other selection tests, which have traditionally determined which students go to the elite colleges.
The change was made mostly to reduce what the government says is a growing education gap in the country: wealthy students go to the best colleges and get the best jobs, keeping the children of poorer families on the social margins. The aim is to reduce the importance of costly tutors and cram schools, partly to help students enjoy a more normal high-school life. But the new system has had the opposite effect. Before, students didn't worry too much about their grade-point averages; the big challenge was beating the standardized tests as high-school seniors. Now students are competing against one another over a three-year period, and every midterm and final test is crucial. Fretful parents are relying even more heavily on tutors and cram schools to help their children succeed.

Parents and kids have sent thousands of angry online letters to the Education Ministry complaining that the new admissions standard is setting students against each other. "One can succeed only when others fail," as one parent said.

Education experts say that South Korea's public secondary-school system is foundering, while private education is thriving. According to critics, the country's high schools are almost uniformly mediocre – the result of an egalitarian government education policy. With the number of elite schools strictly controlled by the government, even the brightest students typically have to settle for ordinary schools in their neighbourhoods, where the curriculum is centred on average students. To make up for the mediocrity, zealous parents send their kids to the expensive cram schools.

Students in affluent southern Seoul neighbourhoods complain that the new system will hurt them the most. Nearly all Korean high schools will be weighted equally in the college-entrance process, and relatively weak students in provincial schools, who may not score well on standardized tests, often compile good grade-point averages.

Some universities, particularly prestigious ones, openly complain that they cannot select the best students under the new system because it eliminates differences among high schools. They've asked for more discretion in picking students by giving
more weight to such screening tools as essay writing or interviews.

President Roh Moo Hyun doesn't like how some colleges are trying to circumvent the new system. He recently criticized "greedy" universities that focus more on finding the best students than faying to "nurture good students". But amid the crossfire between the government and universities, the country's 10th graders are feeling the stress. On online protest sites, some are calling themselves a "cursed generation" and "mice in a lab experiment". It all seems a touch melodramatic, but that's the South Korean school system.

11. According to the passage, the new college-entrance system is designed to ________.

   A. require students to sit for more college-entrance tests
   B. reduce the weight of college-entrance tests
   C. select students on their high school grades only
   D. reduce the number of prospective college applicants

12. What seems to be the effect of introducing the new system?

   A. The system has given equal opportunities to students.
   B. The system has reduced the number of cram schools.
   C. The system has intensified competition among schools.
   D. The system has increased students' study load.

13. According to critics, the popularity of private education is mainly the result of ________.

   A. the government's egalitarian policy
B. insufficient number of schools
C. curriculums of average quality
D. low cost of private education

14. According to the passage, there seems to be disagreement over the adoption of the new system between the following groups EXCEPT

A. between universities and the government
B. between school experts and the government
C. between parents and schools
D. between parents and the government

15. Which of the following adjectives best describes the author's treatment of the topic?

A. Objective.
B. Positive.
C. Negative.
D. Biased.

TEXT B

Wilfred Emmanuel-Jones was a teenager before he saw his first cow in his first field. Born in Jamaica, the 47-year-old grew up in inner-city Birmingham before making a career as a television producer and launching his own marketing agency. But deep down he always nurtured every true Englishman's dream of a rustic life, a dream that his entrepreneurial wealth has allowed him to satisfy. These days he's the owner of a thriving 12-hectare farm in deepest Devon with cattle, sheep and pigs. His latest business venture: pushing his brand of Black Fanner gourmet sausages and
barbecue sauces. "My background may be very urban," says Emmanuel-Jones. "But it has given me a good idea of what other urbanites want."

And of how to sell it. Emmanuel-Jones joins a herd of wealthy fugitives from city life who are bringing a new commercial know-how to British farming. Britain's burgeoning farmers' markets - numbers have doubled to at least 500 in the last five years - swarm with specialty cheesemakers, beekeepers or organic smallholders who are redeploying the business skills they learned in the city. "Everyone in the rural community has to come to terms with the fact that things have changed," says Emmanuel-Jones. "You can produce the best food in the world, but if you don't know how to market it, you are wasting your time. We are helping the traditionalists to move on."

The emergence of the new class of superpeasants reflects some old yearnings. If the British were the first nation to industrialize, they were also the first to head back to the land. "There is this romantic image of the countryside that is particularly English," says Alun Howkins of the University of Sussex, who reckons the population of rural England has been rising since 1911. Migration into rural areas is now running at about 100,000 a year, and the hunger for a taste of the rural life has kept land prices buoyant even as agricultural incomes tumble. About 40 percent of all farmland is now sold to "lifestyle buyers" rather than the dwindling number of traditional farmers, according to the Royal Institution of Chartered Surveyors.

What's new about the latest returnees is their affluence and zeal for the business of producing quality foods, if only at a micro-level. A healthy economy and surging London house prices have helped to ease the escape of the would-be rustics. The media recognize and feed the fantasy. One of the big TV hits of recent years, the "River Cottage" series, chronicled the attempts of a London chef to run his own Dorset farm.

Naturally, the newcomers can't hope to match their City salaries, but many are happy to trade any loss of income for the extra job satisfaction. Who cares if there's
no six-figure annual bonus when the land offers other incalculable compensations?

Besides, the specialist producers can at least depend on a burgeoning market for their products. Today's eco-aware generation loves to seek out authentic ingredients. "People like me may be making a difference in a small way," Jan McCourt, a onetime investment banker now running his own 40-hectare spread in the English Midlands stocked with rare breeds.

Optimists see signs of far-reaching change: Britain isn't catching up with mainland Europe; it's leading the way. "Unlike most other countries, where artisanal food production is being eroded, here it is being recovered," says food writer Matthew Fort. "It may be the mark of the next stage of civilization that we rediscover the desirability of being a peasant." And not an investment banker.

16. Which of the following details of Wilfred Emmanuel-Jones is INCORRECT?

A. He was born and brought up in Birmingham.
B. He used to work in the television industry.
C. He is wealthy, adventurous and aspiring.
D. He is now selling his own quality foods.

17. Most importantly, people like Wilfred have brought to traditional British farming ________.

A. knowledge of farming
B. knowledge of brand names
C. knowledge of lifestyle
D. knowledge of marketing,
18. Which of the following does NOT contribute to the emergence of a new class of farmers?

A. Strong desire for country life.
B. Longing for greater wealth.
C. Influence of TV productions.
D. Enthusiasm for quality food business.

19. What is seen as their additional source of new income?

A. Modern tendency to buy natural foods.
B. Increase in the value of land property.
C. Raising and selling rare live stock.
D. Publicity as a result of media coverage.

20. The sentence in the last paragraph "... Britain isn't catching up with mainland Europe; it's leading the way" implies that ________.

A. Britain has taken a different path to boost economy.
B. more authentic foods are being produced in Britain.
C. the British are heading back to the countryside.
D. the Europeans are showing great interest in country life.

TEXT C

In Barcelona the Catalonians call them castells, but these aren't stereotypical castles in Spain. These castles are made up of human beings, not stone. The people who perform this agile feat of acrobatics are called castellers, and to see their towers take shape is to observe a marvel of human cooperation.
First the castellers form what looks like a gigantic rugby scrummage. They are the foundation blocks of the castle. Behind them, other people press together, forming outward-radiating ramparts of inward-pushing muscle: flying buttresses for the castle. Then sturdy but lighter castellers scramble over the backs of those at the bottom and stand, barefoot, on their shoulders – then still others, each time adding a higher "story".

These human towers can rise higher than small apartment buildings: nine "stories", 35 feet into the air. Then, just when it seems this tower of humanity can't defy gravity any longer, a little kid emerges from the crowd and climbs straight up to the top. Arms extended, the child grins while waving to the cheering crowd far below.

Dressed in their traditional costumes, the castellers seem to epitomize an easier time, before Barcelona became a world metropolis and the Mediterranean's most dynamic city. But when you observe them tip close, in their street clothes, at practice, you see there's nothing easy about what the castellers do—and that they are not merely reenacting an ancient ritual.

None of the castellers can give a logical answer as to why they love doing this. But Victor Luna, 16, touches me on the shoulder and says in English: "We do it because it's beautiful. We do it because we are Catalan."

Barcelona's mother tongue is Catalan, and to understand Barcelona, you must understand two words of Catalan: seny and rauxa. Seny pretty much translates as common sense, or the ability to make money, arrange things, and get things done. Rauxa is reminiscent of our words "raucous" and "ruckus".

What makes the castellers revealing of the city is that they embody rauxa and seny. The idea of a human castle is rauxa—it defies common sense—but to watch one going up is to see seny in action. Success is based on everyone working together to achieve a shared goal.
The success of Carlos Tusquets' bank, Fibanc, shows seny at work in everyday life. The bank started as a family concern and now employs hundreds. Tusquets said it exemplifies how the economy in Barcelona is different.

Entrepreneurial seny demonstrates why Barcelona and Catalonia – the ancient region of which Barcelona is the capital – are distinct from the rest of Spain yet essential to Spain's emergence, after centuries of repression, as a prosperous, democratic European country. Catalonia, with Barcelona as its dynamo, has turned into an economic powerhouse. Making up 6 percent of Spain's territory, with a sixth of its people, it accounts for nearly a quarter of Spain's production – everything from textiles to computers – even though the rest of Spain has been enjoying its own economic miracle.

Hand in hand with seny goes rauxa, and there's no better place to see rauxa in action than on the Ramblas, the venerable, tree-shaded boulevard that, in gentle stages, leads you from the centre of Barcelona down to the port. There are two narrow lanes each way for cars and motorbikes, but it's the wide centre walkway that makes the Ramblas a front-row seat for Barcelona's longest running theatrical event. Plastic armchairs are set out on the sidewalk. Sit in one of them, and an attendant will come and charge you a small fee. Performance artists throng the Ramblas – stilt walkers, witches caked in charcoal dust, Elvis impersonators. But the real stars are the old women and happily playing children, millionaires on motorbikes, and pimps and women who, upon closer inspection, prove not to be.

Aficionados (Fans) of Barcelona love to compare notes: "Last night there was a man standing on the balcony of his hotel room," Mariana Bertagnolli, an Italian photographer, told me. "The balcony was on the second floor. He was naked, and he was talking into a cell phone."

There you have it, Barcelona's essence. The man is naked (rauxa), but he is talking into a cell phone (seny).
21. From the description in the passage, we learn that ________.

A. all Catalonians can perform castells
B. castells require performers to stand on each other
C. people perform castells in different formations
D. in castells people have to push and pull each other

22. According to the passage, the implication of the performance is that ________.

A. the Catalonians are insensible and noisy people
B. the Catalonians show more sense than is expected
C. the Catalonians display paradoxical characteristics
D. the Catalonians think highly of team work

23. The passage cites the following examples EXCEPT ________ to show seny at work.

A. development of a bank
B. dynamic role in economy
C. contribution to national economy
D. comparison with other regions

24. In the last but two paragraph, the Ramblas is described as "a front-row seat for Barcelona's longest running theatrical event". What does it mean?

A. On the Ramblas people can see a greater variety of performances.
B. The Ramblas provides many front seats for the performances.
C. The Ramblas is preferred as an important venue for the events.

D. Theatrical performers like to perform on the Ramblas.

25. What is the main impression of the scenes on the Ramblas?

A. It is bizarre and outlandish.

B. It is of average quality.

C. It is conventional and quiet.

D. It is of professional standard.

TEXT D

The law firm Patrick worked for before he died filed for bankruptcy protection a year after his funeral. After his death, the firm's letterhead properly included him: Patrick S. Lanigan, 1954-1992. He was listed up in the right-hand corner, just above the paralegals. Then the rumors got started and wouldn't stop. Before long, everyone believed he had taken the money and disappeared. After three months, no one on the Gulf Coast believed that he was dead. His name came off the letterhead as the debts piled up.

The remaining partners in the law firm were still together, attached unwillingly at the hip by the bondage of mortgages and the bank notes, back when they were rolling and on the verge of serious wealth. They had been joint defendants in several unwinnable lawsuits; thus the bankruptcy. Since Patrick's departure, they had tried every possible way to divorce one another, but nothing would work. Two were raging alcoholics who drank at the office behind locked doors, but never together. The other two were in recovery, still teetering on the brink of sobriety.

He took their money. Their millions. Money they had already spent long before it arrived, as only lawyers can do. Money for their richly renovated office building in
downtown Biloxi. Money for new homes, yachts, condos in the Caribbean. The money was on the way, approved, the papers signed, orders entered; they could see it, almost touch it when their dead partner – Patrick – snatched it at the last possible second.

He was dead. They buried him on February 11, 1992. They had consoled the widow and put his rotten name on their handsome letterhead. Yet six weeks later, he somehow stole their money.

They had brawled over who was to blame. Charles Bogan, the firm's senior partner and its iron hand, had insisted the money be wired from its source into a new account offshore, and this made sense after some discussion. It was ninety million bucks, a third of which the firm would keep, and it would be impossible to hide that kind of money in Biloxi, population fifty thousand. Someone at the bank would talk. Soon everyone would know. All four vowed secrecy, even as they made plans to display as much of their new wealth as possible. There had even been talk of a firm jet, a six-seater.

So Bogan took his share of the blame. At forty-nine, he was the oldest of the four, and, at the moment, the most stable. He was also responsible for hiring Patrick nine years earlier, and for this he had received no small amount of grief.

Doug Vitrano, the litigator, had made the fateful decision to recommend Patrick as the fifth partner. The other three had agreed, and when Patrick Lanigan was added to the firm name, he had access to virtually every file in the office. Bogan, Rapley, Vitrano, Havarac, and Lanigan, Attorneys and Counselors-at-Law. A large ad in the yellow pages claimed "Specialists in Offshore Injuries." Specialists or not, like most firms they would take almost anything if the fees were lucrative. Lots of secretaries and paralegals. Big overhead, and the strongest political connections on the Coast.

They were all in their mid-to late forties. Havarac had been raised by his father on a shrimp boat. His hands were still proudly calloused, and he dreamed of choking
Patrick until his neck snapped. Rapley was severely depressed and seldom left his home, where he wrote briefs in a dark office in the attic.

26. What happened to the four remaining lawyers after Patrick's disappearance?
   A. They all wanted to divorce their wives.
   B. They were all heavily involved in debts.
   C. They were all recovering from drinking.
   D. They had bought new homes, yachts, etc.

27. Which of the following statements contains a metaphor?
   A. His name came off the letterhead as the debts piled up.
   B. … they could see it, almost touch it when their dead partner...
   C. … , attached unwillingly at the hip by the bondage of mortgages...
   D. … , and for this he had received no small amount of grief.

28. According to the passage, what is the main cause of Patrick stealing the money?
   A. Patrick was made a partner of the firm.
   B. The partners agreed to have the money transferred.
   C. Patrick had access to all the files in the firm.
   D. Bogan decided to hire Patrick nine years earlier.

29. The lawyers were described as being all the following EXCEPT
   A. greedy
B. extravagant
C. quarrelsome
D. bad-tempered

30. Which of the following implies a contrast?

A. … , and it would be impossible to hide that kind of money in Biloxi, population fifty thousand.

B. They had been joint defendants in several unwinnable lawsuits; thus the bankruptcy.

C. There had even been talk of a firm jet, a six-seater.

D. His name came off the letterhead as the debts piled up.
In this section there are four reading passages followed by a total of 20 multiple-choice questions.

Read the passages and then mark your answers on your coloured answer sheet.

TEXT A

We had been wanting to expand our children's horizons by taking them to a place that was unlike anything we'd been exposed to during our travels in Europe and the United States. In thinking about what was possible from Geneva, where we are based, we decided on a trip to Istanbul, a two-hour plane ride from Zurich.

We envisioned the trip as a prelude to more exotic ones, perhaps to New Delhi or Bangkok later this year, but thought our 11- and 13-year-olds needed a first step away from manicured boulevards and pristine monuments.

What we didn't foresee was the reaction of friends, who warned that we were putting our children "in danger," referring vaguely, and most incorrectly, to disease, terrorism or just the unknown. To help us get acquainted with the peculiarities of Istanbul and to give our children a chance to choose what they were particularly interested in seeing, we bought an excellent guidebook and read it thoroughly before leaving.

Friendly warnings didn't change our planning, although we might have more prudently checked with the U.S. State Department's list of troublespots. We didn't see a lot of children among the foreign visitors during our six-day stay in Istanbul, but we found the tourist areas quite safe, very interesting and varied enough even to suit our
son, whose oft-repeated request is that we not see "every single" church and museum in a given city.

Vaccinations weren't needed for the city, but we were concerned about adapting to the water for a short stay. So we used bottled water for drinking and brushing our teeth, a precaution that may seem excessive, but we all stayed healthy.

Taking the advice of a friend, we booked a hotel a 20-minute walk from most of Istanbul's major tourist sites. This not only got us some morning exercise, strolling over the Karakoy Bridge, but took us past a colorful assortment of fishermen, vendors and shoe shiners.

From a teenager and pre-teen's view, Istanbul street life is fascinating since almost everything can be bought outdoors. They were at a good age to spend time wandering the labyrinth of the Spice Bazaar, where shops display mounds of pungent herbs in sacks. Doing this with younger children would be harder simply because the streets are so packed with people; it would be easy to get lost.

For our two, whose buying experience consisted of department stores and shopping mall boutiques, it was amazing to discover that you could bargain over price and perhaps end up with two of something for the price of one. They also learned to figure out the relative value of the Turkish lira, not a small matter with its many zeros.

Being exposed to Islam was an important part of our trip. Visiting the mosques, especially the enormous Blue Mosque, was our first glimpse into how this major religion is practiced. Our children's curiosity already had been piqued by the five daily calls to prayer over loudspeakers in every corner of the city, and the scarves covering the heads of many women.

Navigating meals can be troublesome with children, but a kebab, bought on the street or in restaurants, was unfailingly popular. Since we had decided this trip was not for gourmets, kebabs spared us the agony of trying to find a restaurant each day.
that would suit the adults' desire to try something new amid children's insistence that the food be served immediately. Gradually, we branched out to try some other Turkish specialties.

Although our son had studied Islam briefly, it is impossible to be prepared for every awkward question that might come up, such as during our visits to the Topkapi Sarayi, the Ottoman Sultans' palace. No guides were available so it was do-it-yourself, using our guidebook, which cheated us of a lot of interesting history and anecdotes that a professional guide could provide. Next time, we resolved to make such arrangements in advance.

On this trip, we wandered through the magnificent complex, with its imperial treasures, its courtyards and its harem. The last required a bit of explanation that we would have happily left to a learned third party.

11. The couple chose Istanbul as their holiday destination mainly because

A. the city is not too far away from where they lived.

B. the city is not on the list of the U.S. State Department.

C. the city is between the familiar and the exotic.

D. the city is more familiar than exotic.

12. Which of the following statements is INCORRECT?

A. The family found the city was exactly what they had expected.

B. Their friends were opposed to their holiday plan.

C. They could have been more cautious about bringing kids along.

D. They were a bit cautious about the quality of water in the city.

13. We learn from the couple's shopping experience back home that
A. they were used to bargaining over price.

B. they preferred to buy things outdoors.

C. street markets were their favourite.

D. they preferred fashion and brand names.

14. The last two paragraphs suggest that to visit places of interest in Istanbul

A. guidebooks are very useful.

B. a professional guide is a must.

C. one has to be prepared for questions.

D. one has to make arrangements in advance.

15. The family have seen or visited all the following in Istanbul EXCEPT

A. religious prayers.

B. historical buildings.

C. local-style markets.

D. shopping mall boutiques.

TEXT B

Last month the first baby-boomers turned 60. The bulky generation born between 1946 and 1964 is heading towards retirement. The looming "demographic cliff" will see vast numbers of skilled workers dispatched from the labour force.

The workforce is ageing across the rich world. Within the EU the number of workers aged between 50 and 64 will increase by 25% over the next two decades, while those aged 20-29 will decrease by 20%. In Japan almost 20% of the population is already over 65, the highest share in the world. And in the United States the number
of workers aged 55-64 will have increased by more than half in this decade, at the same time as the 35- to 44-year-olds decline by 10%.

Given that most societies are geared to retirement at around 65, companies have a looming problem of knowledge management, of making sure that the boomers do not leave before they have handed over their expertise along with the office keys and their e-mail address. A survey of human-resources directors by IBM last year concluded: "When the baby-boomer generation retires, many companies will find out too late that a career's worth of experience has walked out the door, leaving insufficient talent to fill in the void."

Some also face a shortage of expertise. In aerospace and defence, for examinationple, as much as 40% of the workforce in some companies will be eligible to retire within the next five years. At the same time, the number of engineering graduates in developed countries is in steep decline.

A few companies are so squeezed that they are already taking exceptional measures. Earlier this year the Los Angeles Times interviewed an enterprising Australian who was staying in Beverly Hills while he tried to persuade locals to emigrate to Toowoomba, Queensland, to work for his engineering company there. Toowoomba today; the rest of the developed world tomorrow?

If you look hard enough, you can find companies that have begun to adapt the workplace to older workers. The AARP, an American association for the over-50s, produces an annual list of the best employers of its members. Health-care firms invariably come near the top because they are one of the industries most in need of skilled labour. Other sectors similarly affected, says the Conference Board, include oil, gas, energy and government.

Near the top of the AARP's latest list comes Deere & Company, a no-nonsense industrial-equipment manufacturer based in Illinois; about 35% of Deere's 46,000 employees are over 50 and a number of them are in their 70s. The tools it uses to
achieve that – flexible working, telecommuting, and so forth - also coincidentally help older workers to extend their working lives. The company spends "a lot of time" on the ergonomics of its factories, making jobs there less tiring, which enables older workers to stay at them for longer.

Likewise, for more than a decade, Toyota, arguably the world's most advanced manufacturer, has adapted its workstations to older workers. The shortage of skilled labour available to the automotive industry has made it unusually keen to recruit older workers. BMW recently set up a factory in Leipzig that expressly set out to employ people over the age of 45. Needs must when the devil drives.

Other firms are polishing their alumni networks. IBM uses its network to recruit retired people for particular projects. Ernst & Young, a professional-services firm, has about 30,000 registered alumni, and about 25% of its "experienced" new recruits are former employees who return after an absence.

But such examinationples are unusual. A survey in America last month by Ernst & Young found that "although corporate America foresees a significant workforce shortage as boomers retire, it is not dealing with the issue." Almost three-quarters of the 1,400 global companies questioned by Deloitte last year said they expected a shortage of salaried staff over the next three to five years. Yet few of them are looking to older workers to fill that shortage; and even fewer are looking to them to fill another gap that has already appeared. Many firms in Europe and America complain that they struggle to find qualified directors for their boards - this when the pool of retired talent from those very same firms is growing by leaps and bounds.

Why are firms not working harder to keep old employees? Part of the reason is that the crunch has been beyond the horizon of most managers. Nor is hanging on to older workers the only way to cope with a falling supply of labour. The participation of developing countries in the world economy has increased the overall supply - whatever the local effect of demographics in the rich countries. A vast amount of work is being sent offshore to such places as China and India and more will go in
future. Some countries, such as Australia, are relaxing their immigration policies to allow much needed skills to come in from abroad. Others will avoid the need for workers by spending money on machinery and automation.

16. According to the passage, the most serious consequence of baby-boomers approaching retirement would be

A. a loss of knowledge and experience to many companies.

B. a decrease in the number of 35- to 44-year-olds.

C. a continuous increase in the number of 50-to 64-year-olds.

D. its impact on the developed world whose workforce is ageing.

17. The following are all the measures that companies have adopted to cope with the ageing workforce EXCEPT

A. making places of work accommodate the needs of older workers.

B. using alumni networks to hire retired former employees.

C. encouraging former employees to work overseas.

D. granting more convenience in working hours to older workers.

18. "The company spends 'a lot of time' on the ergonomics of its factories" (Paragraph Seven) means that

A. the company attaches great importance to the layout of its factories.

B. the company improves the working conditions in its factories.

C. the company attempts to reduce production costs of its factories.

D. the company intends to renovate its factories and update equipment.

19. In the author's opinion American firms are not doing anything to deal with the issue of the ageing workforce mainly because

A. They have not been aware of the problem.
B. They are reluctant to hire older workers.

C. They are not sure of what they should do.

D. They have other options to consider.

20. Which of the following best describes the author's development of argument?

A. introducing the issue---citing ways to deal with the issue---describing the actual status---offering reasons.

B. describing the actual status---introducing the issue---citing ways to deal with the issue---offering reasons.

C. citing ways to deal with the issue---introducing the issue---describing the actual status---offering reasons.

D. describing the actual status---offering reasons---introducing the issue---citing ways to deal with the issue.

TEXT C

(1) The other problem that arises from the employment of women is that of the working wife. It has two aspects: that of the wife who is more of a success than her husband and that of the wife who must rely heavily on her husband for help with domestic tasks. There are various ways in which the impact of the first difficulty can be reduced. Provided that husband and wife are not in the same or directly comparable lines of work, the harsh fact of her greater success can be obscured by a genial conspiracy to reject a purely monetary measure of achievement as intolerably crude. Where there are ranks, it is best if the couple work in different fields so that the husband can find some special reason for the superiority of the lowest figure in his to the most elevated in his wife's.

(2) A problem that affects a much larger number of working wives is the need to re-allocate domestic tasks if there are children. In The Road to Wigan Pier George Orwell wrote of the unemployed of the Lancashire coalfields: "Practically never ... in a working-class home, will you see the man doing a stroke of the housework. 
Unemployment has not changed this convention, which on the face of it seems a little unfair. The man is idle from morning to night but the woman is as busy as ever - more so, indeed, because she has to manage with less money. Yet so far as my experience goes the women do not protest. They feel that a man would lose his manhood if, merely because he was out of work, he developed in a 'Mary Ann'."

(3) It is over the care of young children that this re-allocation of duties becomes really significant. For this, unlike the cooking of fish fingers or the making of beds, is an inescapably time-consuming occupation, and time is what the fully employed wife has no more to spare of than her husband.

(4) The male initiative in courtship is a pretty indiscriminate affair, something that is tried on with any remotely plausible woman who comes within range and, of course, with all degrees of tentativeness. What decides the issue of whether a genuine courtship is going to get under way is the woman's response. If she shows interest the engines of persuasion are set in movement. The truth is that in courtship society gives women the real power while pretending to give it to men.

(5) What does seem clear is that the more men and women are together, at work and away from it, the more the comprehensive amorousness of men towards women will have to go, despite all its past evolutionary services. For it is this that makes inferiority at work abrasive and, more indirectly, makes domestic work seem unmanly, if there is to be an equalizing redistribution of economic and domestic tasks between men and women there must be a compensating redistribution of the erotic initiative. If women will no longer let us beat them they must allow us to join them as the blushing recipients of flowers and chocolates.

21. Paragraph One advises the working wife who is more successful than her husband to

A. work in the same sort of job as her husband.

B. play down her success, making it sound unimportant.
C. stress how much the family gains from her high salary.

D. introduce more labour-saving machinery into the home.

22. Orwell's picture of relations between man and wife in Wigan Pier (Paragraph Two) describes a relationship which the author of the passage

A. thinks is the natural one.

B. wishes to see preserved.

C. believes is fair.

D. is sure must change.

23. Which of the following words is used literally, NOT metaphorically?

A. Abrasive (Paragraph Five).

B. Engines (Paragraph Four).

C. Convention (Paragraph Two).

D. Heavily (Paragraph One).

24. The last paragraph stresses that if women are to hold important jobs, then they must

A. sometimes make the first advances in love.

B. allow men to flirt with many women.

C. stop accepting presents of flowers and chocolates.

D. avoid making their husbands look like "Mary Anns".

25. Which of the following statements is INCORRECT about the present form of courtship?

A. Men are equally serious about courtship.

B. Each man "makes passes" at many women.
C. The woman's reaction decides the fate of courtship.

D. The man leaves himself the opportunity to give up the chase quickly.

TEXT D

From Namche Bazaar, the Sherpa capital at 12,000 feet, the long line threaded south, dropping 2,000 feet to the valley floor, then trudged down the huge Sola-Khumbu canyon until it opened out to the lush but still daunting foothills of Central Nepal.

It was here at Namche that one man broke rank and leaned north, slowly and arduously climbing the steep walls of the natural amphitheater behind the scatter of stone huts, then past Kunde and Khumjong.

Despite wearing a balaclava on his head, he had been frequently recognized by the Tibetans, and treated with the gravest deference and respect. Even among those who knew nothing about him, expressions of surprise lit up their dark, liquid eyes. He was a man not expected to be there.

Not only was his stature substantially greater than that of the diminutive Tibetans, but it was also obvious from his bearing - and his new broadcloak, which covered a much-too-tight army uniform - that he came from a markedly loftier station in life than did the average Tibetan. Among a people virtually bereft of possessions, he had fewer still, consisting solely of a rounded bundle about a foot in diameter slung securely by a cord over his shoulder. The material the bundle was wrapped in was of a rough Tibetan weave, which did not augur that the content was of any greater value - except for the importance he seemed to ascribe to it, never for a moment releasing his grip.

His objective was a tiny huddle of buildings perched halfway up an enormous valley wall across from him, atop a great wooded spur jutting out from the lower lap
of the 22,493-foot Ama Dablum, one of the most majestic mountains on earth. There was situated Tengboche, the most famous Buddhist monastery in the Himalayas, its setting unsurpassed for magnificence anywhere on the planet.

From the top of the spur, one's eyes sweep 12 miles up the stupendous Dudh Kosi canyon to the six-mile-long granite wall of cliff of Nuptse at its head. If Ama Dablum is the Gatekeeper, then the sheer cliff of Nuptse, never less than four miles high, is the Final Protector of the highest and mightiest of them all: Chomolongma, the Mother Goddess of the World, to the Tibetans; Sagarmatha, the Head of the Seas, to the Nepalese; and Everest to the rest of us. And over the great barrier of Nuptse She demurely peaks.

It was late in the afternoon - when the great shadows cast by the colossal mountains were descending into the deep valley floors - before he reached the crest of the spur and shuffled to a stop just past Tengboche's entrance gompa. His chest heaving in the rarefied air, he removed his hand from the bundle--the first time he had done so - and wiped grimy rivulets of sweat from around his eyes with the fingers of his mitted hand.

His narrowed eyes took in the open sweep of the quiet grounds, the pagoda-like monastery itself, and the stone buildings that tumbled down around it like a protective skirt. In the distance the magic light of the magic hour lit up the plume flying off Chomolongma's 29,029-foot-high crest like a bright, welcoming banner.

His breathing calmed, he slowly, stiffly struggled forward and up the rough stone steps to the monastery entrance. There he was greeted with a respectful nameste -"I recognize the divine in you" - from a tall, slim monk of about 35 years, who hastily set aside a twig broom he had been using to sweep the flagstones of the inner courtyard. While he did so, the visitor noticed that the monk was missing the small finger on his left hand. The stranger spoke a few formal words in Tibetan, and then the two disappeared inside.
Early the next morning the emissary - lightened of his load - appeared at the monastery entrance, accompanied by the same monk and the elderly abbot. After a bow of his head, which was returned much more deeply by the two ocher-robed residents, he took his leave. The two solemn monks watched, motionless, until he dipped over the ridge on which the monastery sat, and out of sight.

Then, without a word, they turned and went back inside the monastery.

26. Which of the following words in Paragraph One implies difficulty in walking?
   A. "threaded"
   B. "dropping"
   C. "trudged"
   D. "daunting"

27. In the passage the contrast between the Tibetans and the man is indicated in all the following aspects EXCEPT
   A. clothing
   B. height
   C. social status
   D. personal belongings

28. It can be inferred from the passage that one can get _____ of the region from the monastery.
   A. a narrow view
   B. a hazy view
   C. a distant view
   D. a panoramic view
29. Which of the following details shows that the man became relaxed after he reached the monastery?

A. "...he reached the crest of the spur and shuffled to a stop..."

B. "...he removed his hand from the bundle..."

C. "His narrowed eyes took in the open sweep of the quiet grounds..."

D. "...he slowly, stiffly struggled forward and up the rough stone steps..."

30. From how it is described in the passage the monastery seems to evoke

A. a sense of awe.

B. a sense of piety.

C. a sense of fear.

D. a sense of mystery.
TEST FOR ENGLISH MAJORS (2010)  
-GRADE EIGHT-

PART II READING COMPREHENSION (30 MIN)

In this section there are four reading passages followed by a total of 20 multiple-choice questions. Read the passages and then mark your answers on your coloured answer sheet.

TEXT A

Still, the image of any city has a half-life of many years. (So does its name, officially changed in 2001 from Calcutta to Kolkata, which is closer to what the word sounds like in Bengali. Conversing in English, I never heard anyone call the city anything but Calcutta.) To Westerners, the conveyance most identified with Kolkata is not its modern subway—a facility whose spacious stations have art on the walls and cricket matches on television monitors—but the hand-pulled rickshaw. Stories and films celebrate a primitive-looking cart with high wooden wheels, pulled by someone who looks close to needing the succor of Mother Teresa. For years the government has been talking about eliminating hand-pulled rickshaws on what it calls humanitarian grounds—principally on the ground that, as the mayor of Kolkata has often said, it is offensive to see “one man sweating and straining to pull another man.” But these days politicians also lament the impact of 6,000 hand-pulled rickshaws on a modern city’s traffic and, particularly, on its image. “Westerners try to associate beggars and these rickshaws with the Calcutta landscape, but this is not what Calcutta stands for,” the chief minister of West Bengal, Buddhadeb Bhattacharjee, said in a press conference in 2006. “Our city stands for prosperity and development.” The chief minister—the equivalent of a state governor—went on to announce that hand-pulled rickshaws soon would be banned from the streets of Kolkata.

Rickshaws are not there to haul around tourists. (Actually, I saw almost no tourists in Kolkata, apart from the young backpackers on Sudder Street, in what used
to be a red-light district and is now said to be the single place in the city where the services a rickshaw puller offers may include providing female company to a gentleman for the evening.) It’s the people in the lanes who most regularly use rickshaws—not the poor but people who are just a notch above the poor. They are people who tend to travel short distances, through lanes that are sometimes inaccessible to even the most daring taxi driver. An older woman with marketing to do, for instance, can arrive in a rickshaw, have the rickshaw puller wait until she comes back from various stalls to load her purchases, and then be taken home. People in the lanes use rickshaws as a 24-hour ambulance service. Proprietors of cafés or corner stores send rickshaws to collect their supplies. (One morning I saw a rickshaw puller take on a load of live chickens—tied in pairs by the feet so they could be draped over the shafts and the folded back canopy and even the axle. By the time he trotted off, he was carrying about a hundred upside-down chickens.) The rickshaw pullers told me their steadiest customers are schoolchildren. Middle-class families contract with a puller to take a child to school and pick him up; the puller essentially becomes a family retainer.

From June to September Kolkata can get torrential rains, and its drainage system doesn’t need torrential rain to begin backing up. Residents who favor a touch of hyperbole say that in Kolkata “if a stray cat pees, there’s a flood.” During my stay it once rained for about 48 hours. Entire neighborhoods couldn’t be reached by motorized vehicles, and the newspapers showed pictures of rickshaws being pulled through water that was up to the pullers’ waists. When it’s raining, the normal customer base for rickshaw pullers expands greatly, as does the price of a journey. A writer in Kolkata told me, “When it rains, even the governor takes rickshaws.”

While I was in Kolkata, a magazine called India Today published its annual ranking of Indian states, according to such measurements as prosperity and infrastructure. Among India’s 20 largest states, Bihar finished dead last, as it has for four of the past five years. Bihar, a couple hundred miles north of Kolkata, is where the vast majority of rickshaw pullers come from. Once in Kolkata, they sleep on the street or in their rickshaws or in a dera—a combination garage and repair shop and dormitory managed by someone called a sardar. For sleeping privileges in a dera,
pullers pay 100 rupees (about $2.50) a month, which sounds like a pretty good deal until you’ve visited a dera. They gross between 100 and 150 rupees a day, out of which they have to pay 20 rupees for the use of the rickshaw and an occasional 75 or more for a payoff if a policeman stops them for, say, crossing a street where rickshaws are prohibited. A 2003 study found that rickshaw pullers are near the bottom of Kolkata occupations in income, doing better than only the ragpickers and the beggars. For someone without land or education, that still beats trying to make a living in Bihar.

There are people in Kolkata, particularly educated and politically aware people, who will not ride in a rickshaw, because they are offended by the idea of being pulled by another human being or because they consider it not the sort of thing people of their station do or because they regard the hand-pulled rickshaw as a relic of colonialism. Ironically, some of those people are not enthusiastic about banning rickshaws. The editor of the editorial pages of Kolkata’s Telegraph—Rudrangshu Mukherjee, a former academic who still writes history books—told me, for instance, that he sees humanitarian considerations as coming down on the side of keeping hand-pulled rickshaws on the road. “I refuse to be carried by another human being myself,” he said, “but I question whether we have the right to take away their livelihood.” Rickshaw supporters point out that when it comes to demeaning occupations, rickshaw pullers are hardly unique in Kolkata.

When I asked one rickshaw puller if he thought the government’s plan to rid the city of rickshaws was based on a genuine interest in his welfare, he smiled, with a quick shake of his head—a gesture I interpreted to mean, “If you are so naive as to ask such a question, I will answer it, but it is not worth wasting words on.” Some rickshaw pullers I met were resigned to the imminent end of their livelihood and pin their hopes on being offered something in its place. As migrant workers, they don’t have the political clout enjoyed by, say, Kolkata’s sidewalk hawkers, who, after supposedly being scaled back at the beginning of the modernization drive, still clog the sidewalks, selling absolutely everything—or, as I found during the 48 hours of rain, absolutely everything but umbrellas. “The government was the government of the poor people,” one sardar told me. “Now they shake hands with the capitalists and try to get rid of poor people.”
But others in Kolkata believe that rickshaws will simply be confined more strictly to certain neighborhoods, out of the view of World Bank traffic consultants and California investment delegations—or that they will be allowed to die out naturally as they’re supplanted by more modern conveyances. Buddhadeb Bhattacharjee, after all, is not the first high West Bengal official to say that rickshaws would be off the streets of Kolkata in a matter of months. Similar statements have been made as far back as 1976. The ban decreed by Bhattacharjee has been delayed by a court case and by a widely held belief that some retraining or social security settlement ought to be offered to rickshaw drivers. It may also have been delayed by a quiet reluctance to give up something that has been part of the fabric of the city for more than a century. Kolkata, a resident told me, “has difficulty letting go.” One day a city official handed me a report from the municipal government laying out options for how rickshaw pullers might be rehabilitated.

“Which option has been chosen?” I asked, noting that the report was dated almost exactly a year before my visit.

“That hasn’t been decided,” he said.

“When will it be decided?”

“That hasn’t been decided,” he said.

11. According to the passage, rickshaws are used in Kolkata mainly for the following EXCEPT

A. taking foreign tourists around the city.

B. providing transport to school children.

xxxv
C. carrying store supplies and purchases

D. carrying people over short distances.

12. Which of the following statements best describes the rickshaw pullers from Bihar?

A. They come from a relatively poor area.

B. They are provided with decent accommodation.

C. Their living standards are very low in Kolkata.

D. They are often caught by policemen in the streets.

13. That “For someone without land or education, that still beats trying to make a living in Bihar” (4 paragraph) means that even so,

A. the poor prefer to work and live in Bihar.

B. the poor from Bihar fare better than back home.

C. the poor never try to make a living in Bihar.

D. the poor never seem to resent their life in Kolkata.

14. We can infer from the passage that some educated and politically aware people

A. hold mixed feelings towards rickshaws.

B. strongly support the ban on rickshaws.

C. call for humanitarian actions for rickshaw pullers.

D. keep quiet on the issue of banning rickshaws.

15. Which of the following statements conveys the author’s sense of humor?
A. “…not the poor but people who are just a notch above the poor.” (2 paragraph)

B. “…which sounds like a pretty good deal until you’ve visited a dera.” (4 paragraph)

C. Kolkata, a resident told me, “has difficulty letting go.” (7 paragraph).

D.“…or, as I found during the 48 hours of rain, absolutely everything but umbrellas.” (6 paragraph)

16. The dialogue between the author and the city official at the end of the passage seems to suggest

A. the uncertainty of the court’s decision.

B. the inefficiency of the municipal government.

C. the difficulty of finding a good solution.

D. the slowness in processing options.

TEXT B

Depending on whom you believe, the average American will, over a lifetime, wait in lines for two years (says National Public Radio) or five years (according to customer-loyalty experts).

The crucial word is average, as wealthy Americans routinely avoid lines altogether. Once the most democratic of institutions, lines are rapidly becoming the exclusive province of suckers (people who still believe in and practice waiting in lines). Poor suckers, mostly.

Airports resemble France before the Revolution: first-class passengers enjoy "élite" security lines and priority boarding, and disembark before the unwashed in coach, held at bay by a flight attendant, are allowed to foul the Jetway.

At amusement parks, too, you can now buy your way out of line. This summer I
haplessly watched kids use a $52 Gold Flash Pass to jump the lines at Six Flags New England, and similar systems are in use in most major American theme parks, from Universal Orlando to Walt Disney World, where the have-mores breeze past on their way to their seats.

Flash Pass teaches children a valuable lesson in real-world economics: that the rich are more important than you, especially when it comes to waiting. An NBA player once said to me, with a bemused chuckle of disbelief, that when playing in Canada--get this--"we have to wait in the same customs line as everybody else."

Almost every line can be breached for a price. In several U.S. cities this summer, early arrivers among the early adopters waiting to buy iPhones offered to sell their spots in the lines. On Craigslist, prospective iPhone purchasers offered to pay "waiters" or "placeholders" to wait in line for them outside Apple stores.

Inevitably, some semi-populist politicians have seen the value of sort-of waiting in lines with the ordinary people. This summer Philadelphia mayor John Street waited outside an AT&T store from 3:30 a.m. to 11:30 a.m. before a stand-in from his office literally stood in for the mayor while he conducted official business. And billionaire New York mayor Michael Bloomberg often waits for the subway with his fellow citizens, though he's first driven by motorcade past the stop nearest his house to a station 22 blocks away, where the wait, or at least the ride, is shorter.

As early as elementary school, we're told that jumping the line is an unethical act, which is why so many U.S. lawmakers have framed the immigration debate as a kind of fundamental sin of the school lunch line. Alabama Senator Richard Shelby, to cite just one legislator, said amnesty would allow illegal immigrants "to cut in line ahead of millions of people."

Nothing annoys a national lawmaker more than a person who will not wait in line, unless that line is in front of an elevator at the U.S. Capitol, where Senators and Representatives use private elevators, lest they have to queue with their constituents.
But compromising the integrity of the line is not just antidemocratic, it's out-of-date. There was something about the orderly boarding of Noah's Ark, two by two, that seemed to restore not just civilization but civility during the Great Flood.

How civil was your last flight? Southwest Airlines has first-come, first-served festival seating. But for $5 per flight, an unaffiliated company called BoardFirst.com will secure you a coveted "A" boarding pass when that airline opens for online check-in 24 hours before departure. Thus, the savvy traveler doesn't even wait in line when he or she is online.

Some cultures are not renowned for lining up. Then again, some cultures are too adept at lining up: a citizen of the former Soviet Union would join a queue just so he could get to the head of that queue and see what everyone was queuing for.

And then there is the U.S., where society seems to be cleaving into two groups: Very Important Persons, who don't wait, and Very Impatient Persons, who do--unhappily.

For those of us in the latter group-- consigned to coach, bereft of Flash Pass, too poor or proper to pay a placeholder --what do we do? We do what Vladimir and Estragon did in Waiting for Godot: "We wait. We are bored."

17. What does the following sentence mean? “Once the most democratic of institutions, lines are rapidly becoming the exclusive province of suckers…Poor suckers, mostly.” (2 paragraph)

A. Lines are symbolic of America’s democracy.

B. Lines still give Americans equal opportunities.

C. Lines are now for ordinary Americans only.

D. Lines are for people with democratic spirit only.
18. Which of the following is NOT cited as an examinationple of breaching the line?

A. Going through the customs at a Canadian airport.
B. Using Gold Flash Passes in amusement parks.
C. First-class passenger status at airports.
D. Purchase of a place in a line from a placeholder.

19. We can infer from the passage that politicians (including mayors and Congressmen)

A. prefer to stand in lines with ordinary people.
B. advocate the value of waiting in lines.
C. believe in and practice waiting in lines.
D. exploit waiting in lines for their own good.

20. What is the tone of the passage?

A. Instructive.
B. Humorous.
C. Serious.
D. Teasing.

TEXT C

A bus took him to the West End, where, among the crazy coloured fountains of illumination, shattering the blue dusk with green and crimson fire, he found the café of his choice, a tea-shop that had gone mad and turned. Babylonian, a white palace
with ten thousand lights. It towered above the other building like a citadel, which
indeed it was, the outpost of a new age, perhaps a new civilization, perhaps a new
barbarism; and behind the thin marble front were concrete and steel, just as behind the
careless profusion of luxury were millions of pence, balanced to the last halfpenny.
Somewhere in the background, hidden away, behind the ten thousand lights and acres
of white napery and bewildering glittering rows of teapots, behind the thousand
waitresses and cash-box girls and black-coated floor managers and temperamental
long-haired violinists, behind the mounds of cauldrons of stewed steak, the vanloads
of ices, were a few men who went to work juggling with fractions of a farming, who
knew how many units of electricity it took to finish a steak-and-kidney pudding and
how many minutes and seconds a waitress( five feet four in height and in average
health) would need to carry a tray of given weight from the kitchen life to the table in
the far corner. In short, there was a warm, sensuous, vulgar life flowering in the upper
storeys, and a cold science working in the basement. Such as the gigantic tea-shop
into which Turgis marched, in search not of mere refreshment but of all the
enchantment of unfamiliar luxury. Perhaps he knew in his heart that men have
conquered half the known world, looted whole kingdoms, and never arrived in such
luxury. The place was built for him.

It was built for a great many other people too, and, as usual, they were all there.
It seemed with humanity. The marble entrance hall, piled dizzily with bonbons and
cakes, was as crowded and bustling as a railway station. The gloom and grime of the
streets, the raw air, all November, were at once left behind, forgotten: the atmosphere
inside was golden, tropical, belonging to some high mid-summer of confectionery.
Disdaining the lifts, Turgis, once more excited by the sight, sound, and smell of it all,
climbed the wide staircase until he reached his favourite floor, where an orchestra, led
by a young Jewish violinist with wandering lustrous eyes and a passion for tremolo
effects, acted as a magnet to a thousand girls, scented air, the sensuous clamour of the
strings; and, as he stood hesitating a moment, half dazed, there came, bowing, sleek
grave man, older than he was and far more distinguished than he could ever hope to
be, who murmured deferentially: “For one, sir? This way, please.” Shyly, yet proudly,
Turgis followed him.
21. That “behind the thin marble front were concrete and steel” suggests that

A. modern realistic commercialism existed behind the luxurious appearance.

B. there was a fundamental falseness in the style and the appeal of the café..

C. the architect had made a sensible blend of old and new building materials.

D. the café was based on physical foundations and real economic strength.

22. The following words or phrases are somewhat critical of the tea-shop EXCEPT

A. “…turned Babylonian”.

B. “perhaps a new barbarism’.

C. “acres of white napery”.

D. “balanced to the last halfpenny”.

23. In its context the statement that “ the place was built for him” means that the café was intended to

A. please simple people in a simple way.

B. exploit gullible people like him.

C. satisfy a demand that already existed.

D. provide relaxation for tired young men.

24. Which of the following statements about the second paragraph is NOT true?

A. The café appealed to most senses simultaneously.

B. The café was both full of people and full of warmth.

C. The inside of the café was contrasted with the weather outside.
D. It stressed the commercial determination of the café owners.

25. The following are comparisons made by the author in the second paragraph EXCEPT that

A. the entrance hall is compared to a railway station.
B. the orchestra is compared to a magnet.
C. Turgis welcomed the lift like a conquering soldier.
D. the interior of the café is compared to warm countries.

26. The author’s attitude to the café is

A. fundamentally critical.
B. slightly admiring.
C. quite undecided.
D. completely neutral.

TEXT D

Now elsewhere in the world, Iceland may be spoken of, somewhat breathlessly, as western Europe’s last pristine wilderness. But the environmental awareness that is sweeping the world had bypassed the majority of Icelanders. Certainly they were connected to their land, the way one is complicatedly connected to, or encumbered by, family one can’t do anything about. But the truth is, once you’re off the beaten paths of the low-lying coastal areas where everyone lives, the roads are few, and they’re all bad, so Iceland’s natural wonders have been out of reach and unknown even to its own inhabitants. For them the land has always just been there, something that had to be dealt with and, if possible, exploited—the mind-set being one of land as commodity rather than land as, well, priceless art on the scale of the “Mona Lisa.”
When the opportunity arose in 2003 for the national power company to enter into a 40-year contract with the American aluminum company Alcoa to supply hydroelectric power for a new smelter, those who had been dreaming of something like this for decades jumped at it and never looked back. Iceland may at the moment be one of the world’s richest countries, with a 99 percent literacy rate and long life expectancy. But the project’s advocates, some of them getting on in years, were more emotionally attuned to the country’s century upon century of want, hardship, and colonial servitude to Denmark, which officially had ended only in 1944 and whose psychological imprint remained relatively fresh. For the longest time, life here had meant little more than a sod hut, dark all winter, cold, no hope, children dying left and right, earthquakes, plagues, starvation, volcanoes erupting and destroying all vegetation and livestock, all spirit—a world revolving almost entirely around the welfare of one’s sheep and, later, on how good the cod catch was. In the outlying regions, it still largely does.

Ostensibly, the Alcoa project was intended to save one of these dying regions—the remote and sparsely populated east—where the way of life had steadily declined to a point of desperation and gloom. After fishing quotas were imposed in the early 1980s to protect fish stocks, many individual boat owners sold their allotments or gave them away, fishing rights ended up mostly in the hands of a few companies, and small fishermen were virtually wiped out. Technological advances drained away even more jobs previously done by human hands, and the people were seeing everything they had worked for all their lives turn up worthless and their children move away. With the old way of life doomed, aluminum projects like this one had come to be perceived, wisely or not, as a last chance. “Smelter or death.”

The contract with Alcoa would infuse the region with foreign capital, an estimated 400 jobs, and spin-off service industries. It also was a way for Iceland to develop expertise that potentially could be sold to the rest of the world; diversify an economy historically dependent on fish; and, in an appealing display of Icelandic can-do verve, perhaps even protect all of Iceland, once and for all, from the unpredictability of life itself.
“We have to live,” Halldór Ásgrímsson said in his sad, sonorous voice. Halldór, a former prime minister and longtime member of parliament from the region, was a driving force behind the project. “We have a right to live.”

27. According to the passage, most Icelanders view land as something of

A. environmental value.

B. commercial value.

C. potential value for tourism.

D. great value for livelihood.

28. What is Iceland’s old-aged advocates’ feeling towards the Alcoa project?

A. Iceland is wealthy enough to reject the project.

B. The project would lower life expectancy.

C. The project would cause environmental problems.

D. The project symbolizes an end to the colonial legacies.

29. The disappearance of the old way of life was due to all the following EXCEPT

A. fewer fishing companies.

B. fewer jobs available.

C. migration of young people.

D. imposition of fishing quotas.

30. The 4 paragraph in the passage
A. sums up the main points of the passage.

B. starts to discuss an entirely new point.

C. elaborates on the last part of the 3 paragraph.

D. continues to depict the bleak economic situation.