

## AN INTROSPECTIVE STUDY ON TEST TAKING PROCESS FOR BANKED CLOZE

Gao Xiaoying

PLA Logistics Engineering University

Gu Xiangdong

Chongqing University

### Abstract

The present study investigates the testees test taking process for banked cloze ,focusing on the following two aspects :the information sources and strategies . Employing simultaneous introspection and immediate retrospection ,the study is conducted among 18 non -English major students . After examining their test taking processes ,a framework for analyzing testees test taking processes for banked cloze test is put forward . And then the researcher applies this framework to analyze the testees protocols . Results show that regarding information sources ,clause level information accounts for the largest proportion ,followed by text level information ,sentence level information and extra textual information ,and in terms of strategies ,the testees show a clear preference for bottom up processing strategies ,followed by top down processing strategies and test -wise processing strategies . The verbal protocols have further revealed that the proficient readers seem to have a context based reading model and they prefer to grasp the gist of the text , whereas the less proficient readers often employ word based approaches and they are apt to identify parts of the text .

### Key words

banked cloze ;test taking process ;introspection

### 1 .Introduction

Cloze test is a testing procedure in which the testees are required to restore the letters or words that have been systematically deleted from a continuous text ( Shohamy 1983 ) . Banked cloze test is one of the variants of cloze test . In a banked cloze test ,all the deleted words are put together in a bank and the testees are required to choose the appropriate ones from the bank .

Cloze test has been widely applied in the fields of language testing and teaching at home and abroad . However ,for testing researchers and language teachers ,there has been a long running argument about the construct validity of cloze test ,that is ,what cloze test actually measures and whether it can measure comprehension that ranges beyond the sentence level . Several studies ( e. g. Bachman 1982 ;Brown 1983 ) have found support for the claim that cloze test is capable of measuring testees comprehension ability at text level ,whereas other studies ( e. g. Alderson 1983 ;Markman 1985 ) conclude that cloze test can not measure comprehension beyond the sentence level .

However ,most of these previous studies are product oriented studies . They tend to focus on participants test performance ( i. e. scores ) with little information available on participants cloze test -

---

This paper is a small part of the research project ,“ A Chronological Study on the CET Washback ” ( 07BYY030 ) , supported by National Philosophy and Social Science Foundation of China . The team leader is Gu Xiangdong .

taking process itself. It may not always be the case that the test items measure the ability that the researchers are interested in because testees may use quite different processes to answer the items from those anticipated (Yamashita 2003). Thus the validation of tests involving cognitive processes will not be complete unless it includes some examination of the test taking processes (Storey 1997). In order to reach this goal, recently some researchers (e.g. Feldmann & Stemmer 1987; Storey 1997) began to resort to more direct research methods, introspective methods. Nunan (1992) defines introspection as "the process of observing and reflecting on one's thoughts, feelings, reasoning processes, and mental state with a view to determining the ways in which these processes and states determine our behavior" (p. 115). He uses the term "introspection" to "cover techniques in which data collection is coterminous with the mental events being investigated", as well as "research contexts in which the data are collected retrospectively, that is some time after the mental events themselves have taken place" (Nunan 1992: 115).

Feldmann and Stemmer (1987) use the methods of simultaneous introspection and immediate retrospection to identify the strategies used in the learners' protocols when working on C test. The learners' general problem solving behavior is investigated and specific problem solving strategies are further identified. Seven recall strategies and three evaluation strategies are discovered in the learners' protocols. And then the researchers find that these strategies seem more likely to reflect bottom-up processing and top-down processing. And these kinds of processing are represented as a continuum on which the different strategies approaches are projected.

Storey (1997) employs simultaneous introspection followed by immediate retrospection to investigate EFL learners' processes of taking MC discourse cloze test, which is designed to generate the discourse processing strategies mainly in terms of intersentential comprehension. Four categories of deletions are applied: discourse markers, anaphoric pronouns, lexical substitutes and lexical items. Examining the learners' verbal protocols, the researcher finds that different items entail varying degrees of construct validity. Overall, however, the items are proved to stimulate the students' construct relevant processing, and this test is considered to have a good degree of validity.

Yamashita (2003) investigates how skilled and less skilled readers answer a gap filling test (or rational deletion cloze test) in order to examine whether such tests can measure text level processing ability. Twelve Japanese EFL students (six skilled and six less skilled readers) are asked to provide simultaneous introspection (or think aloud verbal protocols) when they complete a gap filling test. A modified Bachman's (1985) classification system of cloze item types is employed as a basic framework to analyze students' protocols. In her framework, Bachman's (1985) "Across Sentences, Within Text" category is further subdivided into two types, i.e. "Across Sentences, Within Paragraph" and "Across Paragraphs, Within Text". And the categories of "Guessing" and "Missing" are added. As a result, seven categories are used to classify the length of text information required for answering each item in the cloze test. The results of this study show that both skilled and less skilled readers use text level information more frequently than other types of information. The skilled readers, however, use text level information more frequently than the less skilled readers. Overall, the gap filling test generates text level processing and differentiates well between skilled and less skilled readers. This study, therefore, supports the claim that a gap filling test can be used as a test to measure higher order processing ability.

Cheng Xiaokou and Li Shaoshan (2006) conduct a process-oriented study of the construct validity of the MC cloze test in TEM 4 (2003). Employing immediate retrospection, the study is conducted among 20 English majors. Bachman's (1985) classification system of cloze item types is employed as a basic framework to analyze students' protocols. The results show that construct of using redundancy information in the cloze test is well measured and TEM 4 cloze test not only measures test takers' lower-order processing ability (i.e. understating information across clause, within sentence) and also higher-order processing ability (i.e. understanding information across sentence, within text).

But there still exist some limitations in these empirical researches employing introspective methods. These studies either investigate the types of information utilized by testees to complete cloze items (e.g. Yamashita 2003), or explore the strategies employed by the testees (e.g. Feldmann & Stemmer 1987; Storey 1997). Some of them (e.g. Yamashita 2003) even confuse these two elements and involve Guessing in their categorization framework of information sources. The researcher holds that it is more

suitable to recognize Guessing as one of the strategies. And using information sources and employing strategies to answer items are two indispensable elements in cloze test taking process. Therefore, when we do such process-oriented study, both elements are supposed to be involved in one study. Furthermore, the previous empirical studies focus on the cloze types such as the C test (Feldmann & Stemmer 1987), multiple-choice cloze test (Cheng & Li 2006) and gap filling test (Yamashita 2003), however, studies on the test taking processes in banked cloze test is not yet available in the literature.

However, in China, the innovated CET 4, which has been carried out from December 2006, applies Banked Cloze as a new part of Reading Comprehension. As a brand new test format in CET 4, the banked cloze deserves urgently to be investigated. And it is never known that how the testees solve the banked cloze items in their test taking processes. Under these circumstances, the present study, with introspective methods employed, is conducted to explore the testees test taking processes for banked cloze test. It focuses on the following research questions:

- Q1. What information sources do testees use in their processes of taking banked cloze? How do they use these information sources?
- Q2. What strategies do testees employ in their processes of taking banked cloze? How do they employ these strategies?

## 2. Method

### 2.1 Subjects

Eighteen non-English major students in PLA Logistic Engineering University participated in this study. These subjects were grouped according to their reading proficiency as follows: High Group (n = 6), Medium group (n = 6), Low group (n = 6). They were selected and grouped from 212 sophomores of this university according to their scores of the Reading Comprehension Part in CET 4 which they recently took.

### 2.2 Instruments

The banked cloze test used in the present study was chosen from the prototype paper of CET 4 (June of 2006) for its appropriate difficulty level and content. This banked cloze test consists of a text (246 words) in which 10 words have been deleted, and a bank which contains 15 options and is placed after the text. The subjects were required to restore the deleted words through selecting the appropriate options in the bank. The researcher had already made sure that the selected subjects in the present study didn't take this banked cloze test before.

### 2.3 Procedure

The main methodological problem in this study is to decide how to discover the testees test taking processes. Followed by the previous introspective researches (e.g. Feldmann & Stemmer 1987; Storey 1997), the methods of simultaneous introspection and immediate retrospection were chosen in the present study. In terms of simultaneous introspection which was also known as concurrent talking or thinking aloud or verbalization of specific cognitions, verbal protocols were generated at the same time as the individual was working on the task (Faerch & Kasper 1987). Regarding immediate retrospection, verbal protocols were produced immediately after completion of the task while traces of the original cognition were still present in short term memory (Faerch & Kasper 1987). The experimental procedures were carried out in the following phases.

The subjects were firstly orally informed about the objectives of the study and were then trained in how to provide simultaneous introspection protocols by using verbal arithmetic (Ericsson & Simon 1984; Green 1998). And then a trial banked cloze test was given to them for the purpose of making sure whether they have understood everything correctly and familiarizing them with the task.

Subsequent to the training phase, the subjects were asked to proceed to the simultaneous introspection phase one by one. Each subject was left alone in a quiet room and asked to provide simultaneous introspection either in Chinese or English during their working on the banked cloze. Subjects were prompted to continue talking by the researcher if they fell silent for a period of time. No time limit was set

for the subjects to complete the test with introspection. All the subjects' utterances were audio taped.

And then in the immediate retrospection phase, the subject and researcher listened to the audio tape together immediately after the subject had completed the test. The subject could spontaneously comment on his or her utterances during the previous test phase and answer the researcher's questions concerning the utterances. It is intended to provide additional information about test-taking process for banked cloze test.

## 2.4 Data Analysis

The verbal protocol data obtained from the testees covered all the processes of taking the banked cloze test, including general reading processes, item answering processes, and retrospective interview. For the purposes of analysis, the recorded verbal protocols were divided into idea units consisting of one clause (i.e. subject, verb, and modifiers). Each idea unit was examined in terms of information sources and strategies in order to answer the two research questions.

First, each idea unit was labeled in terms of information sources and strategies. Second, the information source or strategy, which occurred at least 5 times in the testees' protocols, was given a descriptive name. Third, three testees' protocols were randomly selected and another rater was asked to independently categorize the protocols in terms of information sources and strategies. After that, his classifications were compared with those of the researcher and the percent of inter-rater agreement was calculated as 86 percent, indicating that the coding of the verbal protocols is reliable.

## 3. Findings and Discussions

The research questions of the present study focus on two main aspects, i.e. information sources and strategies. Therefore the data will be analyzed and discussed mainly from these two aspects. After examining the protocols obtained from the testees, a framework for analyzing test-taking for banked cloze is discovered (see Figure 1). Four categories of information sources have been revealed, including clause level, sentence level, text level and extra-textual information. Furthermore, twelve types of strategies are also found in the protocols and they can be further classified into three categories, top-down processing strategies, bottom-up processing strategies and test-wise processing strategies.

Figure 1. A framework for analyzing testees' test-taking processes for banked cloze test

The further discussion and explanation on each information source and strategy in the framework will be carried out in the following parts . Before that ,some words of caution are necessary . Although we have listed all the information and strategies so far identified in these protocols ,the list is ,of course ,an open one and will be extended and modified as necessary .

### 3 .1 Information Sources

All the testees protocols of solving the cloze items are classified and the frequency and proportion of each information category is calculated and presented according to the level of the testees reading proficiency and the correct or incorrect responses to the items . In doing so ,the researcher intends to reveal how these categories of information work in the testees test taking process . The results are shown in Table 1 .

Table 1 . Frequencies and proportions of information categories used by three groups of testees

Categories	Correct or incorrect	High		Medium		Low		Total	
		F(n )	P( %)	F(n )	P( %)	F(n )	P( %)	F(n )	P( %)
Clause	Correct	37	12 .6	26	8 .9	13	4 .4	76	25 .9
	Incorrect	19	6 .5	24	8 .2	19	6 .5	62	21 .2
Sentence	Correct	19	6 .5	13	4 .4	10	3 .4	42	14 .3
	Incorrect	3	1 .0	9	3 .1	11	3 .8	23	7 .8
Text	Correct	12	4 .1	20	6 .8	10	3 .4	42	14 .3
	Incorrect	8	2 .7	14	4 .8	3	1 .0	25	8 .5
Extra textual	Correct	7	2 .4	2	0 .7	5	1 .7	14	4 .8
	Incorrect	3	1 .0	1	0 .3	5	1 .7	9	3 .1
Total	Correct	75	25 .6	61	20 .8	38	13 .0	174	59 .4
	Incorrect	33	11 .3	48	16 .4	38	13 .0	119	40 .6

As is shown in Table 1 ,the total information sources are supposed to be 180 ,for there are 18 testees and 10 items in the test . However ,the number of the information sources identified in the present study is counted to be 293 (174 +119 ) .In the light of this result ,it seems reasonable to suggest that the testees tend to use two or three information sources to answer one item . This result seems to reflect the interactive nature of reading processes in which various information sources interact with each other to help readers construct a meaning representation ( Carr & Levy 1990 ) . And Table 1 also displays that the amounts of information sources used by the High Group and Low Group are counted to be 108 (75 +33 ) and 79 (38 +38 ) respectively . In order to examine whether there is a significant difference between the amount of information sources used by the two groups ,Chi-Square test is conducted here ,which is a nonparametric procedure to test frequencies when the total number of observations is more than 40 and the expected frequencies of all the cells are more than 5 . The significance level is set at  $p < 0 .05$  ,which will be applied in the following analyses . The result obtained by Chi-Square test is that the significance is 0 .003 ,and less than 0 .05 ,suggesting that the total number of information sources used by the High Group is significantly larger than that of Low Group . This result indicates that the proficient readers are more active in their attempt to comprehend the text and most of their answers are induced by more than one information source . Not surprisingly ,the High Group performed more successfully than Low group in almost all the categories ( with 25 .6 % and 13 % respectively ) . This is also presumably due to the proficient readers better understanding of the text ,which makes it easier for them to find various kinds of and combinations of information sources in order to solve the cloze items .

#### 1 ) Clause level Information

Clause level information refers to the information provided by the clause in which an item appears . Of the four categories of information ,clause level information is utilized by the testees most frequently (see Table 1 ) . This result is not expected by the researcher for Yamashita (2003 )s research on gap filling test had claimed that cognitive processes requiring text level information are generally provoked more

than any other type of information. Accordingly, we make the further exploration on the cases of applying the clause-level information in order to find why the testees often utilize the clause-level information as the clues to answer items and how they use this category of information.

As a result, a common pattern is observed in the testees' protocols. They often determine the part of speech of the deleted word and then choose the appropriate answer from the options with such part of speech. It is hypothesized that this behavioral pattern found in the testees' test-taking processes in the banked cloze test may be attributed to the effects of such test method, as such behavioral pattern is not identified in the Yamashita's study on the gap-filling test and Chen Xiaokou and Li Shaoshan's (2006) study on the MC cloze test. When asked why they often determine the part of speech of the deleted word, many testees report that the answer can be reached easier and faster by doing so. In a banked cloze test, the options are put together in one bank, which offers an opportunity for the testees to classify the options according to their parts of speech and then choose the answer from the options with the same part of speech as that of the deleted word. This common pattern observed in the testees' protocols may expound the testees' frequent use of the clause-level information sources. According to Bachman's (1990) test performance is affected by the characteristics of the methods used to elicit test performance and it is important not only to understand the nature and extent of these effects, but also to control or minimize them. Therefore, as for designing the options in the banked cloze test, it is suggested to create enough distractors with the same part of speech as that of the correct answer.

As is shown in Table 1, with the help of the clause-level information the testees of High Group can get more correct answers (12.6%) than Low group (4.4%). It seems that the proficient readers answer the banked cloze test more successfully with the help of clause-level information than the less proficient readers. Different processing behaviors of the two groups can be observed in some protocols of the testees, which may provide one possible explanation for this finding. The testees of High Group tend to use clause-level information as a supplementary source to get the answers. For instance, after reaching the answer with the text-level information, the testee of High Group in Example 1 turned to the clause-level information to confirm his choice (see Example 1). But when working on the same item, the testee of Low Group employed clause-level information as the sole source to generate the answer (see Example 2).

Example 1 (a testee of the High Group, working on item 6)

Brought the most, most. That should be the worst weather because the whole text is talking about something bad and the numbers in the following sentence show that it has caused great damage. And the answer should be an adjective. So I chose J, "destructive".

Example 2 (a testee of the Low Group, working on item 6)

Brought the most. What weather. Describe the weather. An adjective. "Deliberately" weather, what's meaning, it is an adverb. "Tropical" and "notify" are adjective. Worth is a noun. It's "notify". (His retrospective protocols: I presumed that the answer is either "tropical" or "notify" and then I chose "notify" at random.)

[The original text: The 1982-83 El Niño brought the most 6 weather in modern history.] [The Key: destructive]

## 2) Sentence-level information

Sentence-level information refers to the information provided by a larger context than the clause in which an item appears, but within the sentence. There are 42 cases (with proportion of 14.3%) in which the testees get the correct answers by employing the sentence-level information (see Table 1). This result is consistent with that of the previous researches (e.g. Bachman 1982; Chen & Li 2006). All of them maintain that processing ability in the sentence level information can be tapped in the cloze test-taking process.

It can also be inferred from the statistics of Table 1 that 86.4% (19 out of 22) of sentence-level information used by the High Group can help to reach the correct answers while only 47.6% (10 out of

21 ) of such information can be used correctly by the Low group . Further examination on the protocols of these groups provides some possible explanations for this phenomenon . A behavioral pattern frequently spotted among the testees of High Group is that they tend to take the sentence as one unit and do not make the closure until they finish the whole sentence . However , the testees of the Low Group often make haste to seek the answer from the options whenever they meet the blank . This behavioral pattern can be best illustrated by the following two examples :

Example 3 ( a testee of the High Group , working on item 9 )

Nowadays weather experts can forecast when an EI Nino will , will come , will . A verb should be filled in this blank . But they are still not , not completely sure what leads to it or what affects how strong it will be . Will , a verb , come , it should be "strike " .

Example 4 ( a testee of the Low Group , working on item 9 )

Nowadays weather experts are able to forecast . Forecast is supposed to have the meaning of find . When an EI Nino will , will come or happen . Yes , "stable " . I choose "stable " .

[ The original text : Nowadays , weather experts are able to forecast when an EINino will 9 , but they are still not 10 sure what leads to it or what affects how strong it will be . ] [ The Key : strike ]

The above protocols reveal that although the testee of High Group in example 3 had inferred the deleted word might have the meaning of "come " , he still tried to complete the whole sentence . On the contrary , the testee of Low Group in example 4 made choice of "stable " immediately after he got some idea of the deleted word .

### 3 ) Text level information

Text level information refers to the information provided by a larger context than the sentence in which an item appears , but a context from within the text . As has been mentioned above , the most frequently discussed topic in the cloze researches is whether the cloze test can measure comprehension that ranges beyond the sentence level . And the previous researchers never reach a consensus on it . One of the main purposes of the present study is to try to brush off this controversy by applying introspective methods to look into the mental processes of the testees when they take the cloze . And it is expected to discover whether the testees applied the text level information during their process of taking the banked cloze test and how they use such category of information to answer the test .

As is shown in Table 1 , the text level information is the second most frequent category to be invoked when the testees take the banked cloze , with 42 cases ( 14 . 3 % ) to reach the correct answers . It implies that testees have to utilize text level information frequently in order to complete the banked cloze test . This result lends support to the conclusions of the previous researches ( e . g . Bachman 1985 ; Chavez -Oller et al . 1994 ; Yamashita 2003 ) that cloze tests are sensitive to intersentential or text level constraints and can be used as a procedure to measure global level comprehension .

A closer inspection of the protocols shows that some individual differences exist between the High Group and Low Group in their application of text level information . Some testees of High Group reach their answers quickly by using text level information ( see Example 5 ) , whereas the Low Group testees seem to have used various information sources gradually in their attempts to get their answers ( see Example 6 ) .

Example 5 ( a testee of the High Group , working on item 1 )

Change in the climate of the world . This strange , happens every five to eight years . Its "phenomenon " . ( Her retrospective protocol : I didn't get the answer when I read the passage for the first time . But when I read it again I realized that "phenomenon " should be filled in this blank because the whole passage talks about the natural disaster . )

[ The original text : This strange 1 happens every five to eight years . ] [ The Key : phenomenon ]

Example 6 (a testee of the Low Group ,working on item 4 )

The rainfall is increased across South America . The blank is within this sentence . Flood . The deleted word should be related to “flood ” . And it is something that can cause the damage since the whole passage talks about damage . And also it is said that it has far reaching effect in the first sentence of this paragraph , which indicates that it has great influence . The deleted word should be an adjective . “Destructive ” has been selected . There are only two adjectives left , “tropical ” and “stable ” . It seems that “stable ” has the similar meaning with “destructive ” . I don't know the meaning of “tropical ” and usually I won't select the unknown word in the test . So I choose “stable ” .

[ The original text : The rainfall is increased across South America , 4 floods to Peru . ] [ The Key : bringing ]

The proficient reader in Example 5 reached the answer decidedly just with the help of text level information . However , when the less proficient reader in Example 6 answered the item , he firstly utilized the clause level information to identify the sentence structure , and then used the text level information to infer the meaning of the deleted word . Moreover he judged the part of speech of the deleted word with the help of the clause level information . Finally he employed such extra textual information as his test experience to make a decision .

#### 4 ) Extra textual information

Extra textual information refers to information not provided by the text . There are 14 cases ( with the proportion of 4 .8 % ) in which the testees apply the extra textual information to reach the correct answers ( see Table 1 ) . The protocols reveal that the extra textual information utilized by the testees in the present study derives from two sources : their background knowledge and their test taking experience .

Background knowledge , it is also called prior knowledge of content in tests of reading comprehension ( Bachman 1990 ) . Many previous researches , such as Erickson and Molloy ( 1983 ) , Alderson and Urquhart ( 1983 ) , have provided quite convincing evidence of an interaction between testees familiarity with content area and performance on tests of listening and reading comprehension ( as cited in Bachman 1990 : 273 ) . The findings in the present study concur with those of the previous studies . It can be illustrated by the following example .

Example 7 ( a testee of the High Group , working on Item 9 )

Weather experts are able to forecast when an EI Nino will , will , come , but they are not sure what leads to it or what affects how strong it will be . But there is not “come ” in the bank . So it should be “strike ” . ( His retrospective protocol : The whole text is mainly talking about a wind , storm wind . And according to my general knowledge , it is often said that the storm wind strikes somewhere . )

[ The original text : Nowadays , weather experts are able to forecast when an EI Nino will 9 , but they are still not 10 sure what leads to it or what affects how strong it will be . ] [ The Key : strike ]

In addition to the background knowledge , the testees test taking experience is also applied as the extra textual information when they are working on the items . For instance , in Example 8 the testee inferred the meaning of the deleted word with sentence level information and then used his previous test-taking experience , i. e. “attraction ” is the vocabulary frequently tested , as a supplementary source to reach the answer .

Example 8 ( a testee of the High Group , working on Item 2 )

The temperatures of ocean rise . “As ” here means “when ” . It seems that when the wind comes together , the ocean temperature will rise as a result . “Attraction ” means “exciting the interests ” as well as “causing to come together ” . Furthermore , “attraction ” is often tested in the examination which I have taken . I always encounter this vocabulary in the test . So I chose “attraction ” .

[ The original text : As the trade winds lessen in 2 , the ocean temperatures rise , causing the Peru

current flowing in from the east to warm up by as much as 5 .] [The Key strength]

### 3.2 Strategies

Strategies are defined deliberate cognitive steps that learners can take to assist in acquiring, storing, and retrieving view information (Paris 1983, as cited in Anderson 1991). After examining the protocols of the testees, twelve strategies are discovered. We make the further categorization of these twelve strategies into three groups. Following Feldmann and Stemmer analysis (1987), we combine eight of the twelve strategies into two categories: top-down processing strategies (i.e. previewing the text before answering the items, evaluating critically what is read, predicting or guessing the meaning of the text, asking oneself questions) and bottom-up processing strategies (i.e. guessing the meanings of unknown words, stating failure to understand unknown words, summarizing the information of the text, analyzing the syntax). The four remaining strategies (i.e. reading the options before reading the text, eliminating the options, matching the blank of an item with the options, guessing blindly) are labeled as test-wise processing strategies. The frequencies and proportions of strategies in each category used by three groups of testees are calculated as follows:

Table 2. Frequencies and proportions of strategy categories used by three groups of testees

Categories of strategies	High		Medium		Low		Total	
	F(n)	P(%)	F(n)	P(%)	F(n)	P(%)	F(n)	P(%)
Top-down Processing Strategies	39	12.3	28	8.9	20	6.3	87	27.5
Bottom-up Processing Strategies	62	19.6	60	19.0	38	12.0	160	50.6
Test-wise Processing Strategies	21	6.6	22	7.0	26	8.2	69	21.8
Total	122	38.6	110	34.8	84	26.6	316	100

Table 2 displays that there are 316 protocols which are reported to use strategies. The testees of High Group employ strategies 122 times while those of Low Group employ 84 times. The Chi-Square analysis shows that the frequency of the strategies employed by High Group is significantly higher than that of Low Group ( $\chi^2 = 14.019, p = 0.000$ ), indicating that the proficient readers tend to employ more strategies in their test-taking processes than the less proficient readers. And the testees of High Group can also use more types of strategies than those of Low Group since the High Group employs twelve types of strategies and the Low Group only uses nine. Evaluating critically what is read, asking oneself questions and summarizing the information of the text are three strategies which do not occur in the protocols of Low Group. These findings reveal that the positive relationship exists between the testees' reading proficiency and the frequency of their used strategies. The proficient readers use strategies more often than less proficient readers. Reading is an active process in which the reader makes efficient use of strategies to understand the text (Olshavsky 1976). In the light of this result, it seems reasonable to suggest that the proficient readers are more active in their attempt to comprehend than less proficient ones.

Of three categories of strategies, testees show a clear preference for bottom-up processing strategies, followed by top-down processing strategies and test-wise processing strategies. Bottom-up processing strategies are applied most frequently by the testees, with 160 cases found in the protocols (see Table 2). And difference can be found in the frequencies of the top-down processing strategies used by the testees of High Group and Low group, with High Group 39 cases (12.3%) and Low group 20 cases (6.3%) respectively (see Table 2). The proficient readers tended to employ top-down processing strategies more frequently than less proficient readers. Finally, test-wise processing strategies occurred 69 times (21.8%) (see Table 2) in the testees' protocols. They were mainly employed when the testees failed to answer items through the information of text. More detailed analyses of these three categories of strategies will be carried out in the following section.

1 ) Top-down processing strategies

Top-down reading models suggest that processing of a text begins in the mind of the readers with meaning-driven processes, or an assumption about the meaning of a text (Dechant 1991). From this perspective, readers identify letters and words only to confirm their assumptions about the meaning of the text (Dechant 1991). Four top-down strategies have been discovered in the testees' top-down processing. Table 3 provides the results of the further examination on these four top-down strategies.

Table 3. Frequencies of top-down processing strategies used by three groups of testees

Top-down Strategies	High	Medium	Low	Total
Preview the text before answering the items	5	4	1	10
Evaluate critically what is read	5	5	0	10
Predict or guess the meaning of the text	25	18	19	62
Ask oneself questions	4	1	0	5

Of these four strategies, predicting or guessing the meaning of the text occurs most frequently when testees process the text (see Table 3). The testees' performance in the present study supports Goodman's statement (1982) that reading is a psycholinguistic guessing game, in which readers guess or predict the text's meaning on the basis of textual information, and their existing, activated, knowledge (as cited in Alderson, 2003:17). The format of the banked cloze test has actually activated such psycholinguistic guessing game since some words are deleted from the text and replaced with blanks. Many testees try to guess the deleted word whenever they encounter a blank (see Example 9).

Example 9 (a testee of the High Group, working on Item 6)

EI Nino brought the most, most awful, or worse weather. (Her retrospective protocols: It seemed that the whole text was taking about a bad thing, and the statistics in the sentence following this item were very large, which implied that it had caused great damage.)

[The original text: The 1982-83 EI Nino brought the most 6 weather in modern history.] [The Key: destructive]

Previewing the text before answering the items is the second frequently used top-down strategy by the testees. And the testees of High Group have higher frequent use of this strategy. As is shown in Table 3, five out of six High Group readers chose to preview the whole text before they worked on the items. On the contrary, only one out of six Low group readers tended to preview the text. One possible explanation for this result is that proficient readers incline to answer the item only after they have possessed an overall understanding of the text. They want to grasp the gist of what the author is trying to say regardless of the details. However, when less proficient readers process the text, they might perceive that it resemble a jigsaw puzzle and just identify parts of the text. Therefore they often complete the items without previewing the text.

Within these four strategies, it is especially worth highlighting that none of testees in Low group report to use such strategies as evaluating critically what is read and asking oneself questions. However, the testees of High Group apply the strategy of critically evaluating 5 times and asking oneself questions 4 times (see Table 3). Proficient readers' higher frequent use of these two strategies may imply that they understand the text more actively than less proficient readers.

2 ) Bottom-up processing strategies

Bottom-up models operate on the principle that the written text is hierarchically organized (i.e. on the grapho-phonetic, phonemic, syllabic, morphemic, word, and sentence levels) and that the reader first processes the smallest linguistic unit, gradually compiling the smaller units to decipher and comprehend the higher units (e.g. sentence syntax) (Dechant 1991). A bottom-up reading model emphasizes a single-

direction ,part to whole processing of a text .Itis also known as part to whole model . Four strategies are discovered in the testees bottom up reading models (see Table 4 ) .

Table 4 . Frequencies of bottom up processing strategies used by three groups of testees

Bottom up Strategies	High	Medium	Low	Total
Guess the meanings of unknown words	8	6	2	16
State failure to understand unknown words	3	5	13	21
Sum marize the information of the text	4	2	0	6
Analyze the syntax	47	47	23	117

Of the four bottom up processing strategies ,analyzing the syntax is most frequently used by the testees (see Table 4 ) . Many testees tend to analyze the sentence structure to identify part of speech of the deleted word and then seek the option with such part of speech when they answer the items . And the findings in Table 4 reveal that there is greater frequency of such strategy in High Group than in Low group . It may be explained by the reason thattestees of High Group have betterlinguistic knowledge than those of Low group .

Some differences can be found when testees with different levels of reading proficiency encounter unknown words . The results in Table 4 show that the testees of Low group state failure to understand unknown words 13 times ,while testees of High Group state only 3 times . The testees of High Group seem to allocate their attention selectively ,probably realizing that they do not need to understand all the words in a sentence ,or perhaps avoiding pronouncing an unfamiliar word . Some of High Group testees incline to guess the meaning of unknown words which is employed 8 times ,whereas testees of Low group use this strategy only 2 times (see Table 4 ) . Less proficient readers tend to read sentence by sentence and thus experience more lexical problems . When they don t understand words ,they feel they can t understand the sentence and seem to feel totally defeated . However ,when proficient readers have difficulty in understanding a word ,they use various and combined approaches to solve their vocabulary problem ,such as resorting to textual information ,fleshing it out ,making it concrete . As a result ,their vocabulary problems do not interfere their on going reading . The emphasis of the less proficient readers on identifying unknown words suggests their word level model of reading . The proficient readers seem to have a more meaning based approach and don t worry about the meanings of words if they could extract the gist of the sentence .

None of Low group testees employ the strategy of sum marizing the information of the text ,while the testees of High Group do use this strategy in their test taking processes (see Table 4 ) . It supplies additional evidence that less proficient readers might perceive the text as a jigsaw puzzle and they often identify parts of the text when they process the text . Proficient readers are more interested in constructing overall comprehension of the text ,as is shown in the following example .

Example 10 (a testee of the High Group )

Nowadays ,weather experts are able to forecast when an EI Nino will ,but they are still not ,sure what leads to it . This text is talking about winds . Now begin to answer the items .

3 ) Test wise processing strategies

Test wise processing strategies is defined as testee s capacity to utilize the characteristics and formats of the test and /or test taking situation to receive a high score ( Crehan , Koehler & Slakter 1974 ) . Examination of the protocols reveals that four test wise processing strategies are exploited when the testees take the banked cloze test .

Table 5 . Frequencies of test-wise processing strategies used by three groups of testees

Test-wise processing Strategies	High	Medium	Low	Total
Read the options before reading the text	4	4	3	11
Eliminate the options	8	7	4	19
Matching the blank of an item with the options	5	10	8	23
Guess blindly	4	1	11	16

Matching the blank of item with the options is the test-wise processing strategy most frequently employed during answering the banked cloze items (see Table 5). When using the strategy of matching the blank of an item with the options, the testees often put the given options one by one in the blanks in order to see which is suitable in the context. Sometimes the testees don't match the blanks with all of the options. They firstly identify the part of speech of the deleted word in the blank and then match this blank with the options with such part of speech (see Example 11).

Example 11 (a testee of the High Group, working on Item 3)

The warming of the ocean has far-reaching effects. The hot, humid air over the ocean causes severe, severe, thunderstorms. The blank should be filled with an adjective. "Stable", "tropical", "destructive". It is caused by the hot and humid air. So "tropical" is more suitable.

[The original text: The hot, humid (潮湿的) air over the ocean caused severe 3 thunderstorms.] [The Key: tropical]

The second frequently used test-wise processing strategy is eliminating the options. Sometimes the testees select an answer not because it is thought to be correct, but because the other options do not seem reasonable or are not understandable. Such strategy is named eliminating the options. There is an example for this: When the testee of Example 12 read the text for the first time, he could not figure out Item 5 so he skipped it. When he came back to Item 5 again, there were only three options left which were "starvation", "phenomenon" and "exhaustion". And then he reported as follows:

Example 12 (a testee of the Low Group, working on item 5)

I seldom encounter starvation and exhaustion according to my test experiences, and I will not choose the words which I am not familiar. phenomenon is more familiar to me than the other two options. So I choose phenomenon although I'm not very sure what it means.

[The original text: So while some parts of the world prepare for heavy rains and floods, other parts face drought, poor crops and 5] [The Key: starvation]

Guessing blindly is labeled to describe the performance that testees randomly choose one from the given options without any particular consideration. Example 13 serves as a good illustration for the use of this strategy. It is observed that this testee just made the random choice of "strike" and not surprisingly he didn't choose the correct answer.

Example 13 (a testee of the Low Group, working on Item 4)

The rainfall is 0 across South America. I don't know how to answer this item. I choose 0, "strike". (His retrospective protocol: I just chose it at random.)

[The original text: The rainfall is increased across South America, 4 floods to Peru.] [The Key: bringing]

There are also 11 cases of reading the options before reading the text. The testees might employ this strategy to make them familiar with the options and keep the options in mind when they answer the items. A common pattern can be found when the testees read the options. They tend to identify the meaning and part of speech of each option. Some of them even take some notes on the test paper.

## 4 . Conclusions

The present study has investigated the testees processes of taking a banked cloze test . After examining their verbal protocols reported simultaneously when they are taking the test and immediately after that ,a framework for analyzing testees processes of taking the banked cloze test is proposed , which contains two aspects :information sources and strategies .

In terms of information sources ,the information they use derives from clause ,text ,sentence ,and extra text . Of these four categories ,clause level information occupies the largest proportion ,followed by text level ,sentence level ,and extra textual information . The reason why the testees frequently apply clause level information lies in the fact that they often use the syntax clues to determine the parts of speech of the deleted words . Text level information is second most frequently used in the testees test-taking processes ,which implies that banked cloze test can measure comprehension that ranges beyond the sentence level .

Regarding strategies ,strategies identified through examining testees verbal protocols include predicting or guessing the meaning of the text ,asking oneself questions ,previewing the text before answering the items ,evaluating critically what is read ,analyzing the syntax ,stating failure to understand unknown words ,guessing the meanings of unknown words ,summarizing the information of the text ,reading the options before reading the text ,eliminating the options ,matching the blank of an item with the options ,and guessing blindly . And these twelve strategies can be further classified into three categories :top down processing strategies ,bottom up processing strategies and test-wise processing strategies . The testees show a clear preference to bottom up processing strategies ,followed by top down processing strategies and test-wise processing strategies . In addition ,proficient readers seem to have a context level model of reading and they are looking for the gist of what the author is trying to say regardless of the details ,whereas less proficient readers often employ word based approaches ,they might perceive the text resemble a jigsaw puzzle ,and only parts of text can be identified when they process the text .

Finally ,there are several limitations in the study which call for cautions in interpreting the results . Firstly ,because of the time consuming nature of the introspective methodology ,this study has only been conducted among eighteen subjects and with one banked cloze test . Therefore ,it should be careful to generalize the results to a wider range . The study should be replicated and the results should be confirmed by other studies with a greater variety of subjects and cloze items . In addition ,the present study does not set a time limit for the testees for fear of influencing their introspective procedure . However ,the condition of no time limit is different from normal test taking situations .

## References

- Alderson ,J . C . 1979 . The cloze procedure and proficiency in English as a foreign language . *TESOL Quarterly* 13 /2 :219 -227 .
- Alderson ,J . C . 1983 . The cloze procedure and proficiency in English as a foreign language . In J . W . Oller (ed .) . *Issues in Language Testing Research* ,pp . 205 -217 . Rowley :Newbury House .
- Alderson ,J . C . 2003 . *Assessing Reading* . Cambridge :CUP .
- Anderson ,N . J . 1991 . Individual differences in strategy use in second language reading and testing . *Modern Language Journal* 75 /4 :460 -472 .
- Bach man ,L . F . 1982 . The trait structure of cloze test scores . *TESOL Quarterly* 16 /1 :61 -70 .
- Bach man ,L . F . 1985 . Performance on cloze tests with fixed ratio and rational deletions . *TESOL Quarterly* 19 /3 :535 -556 .
- Bach man ,L . F . 1990 . *Fundamental Considerations in Language Testing* . Oxford :OUP .
- Brown ,J . D . 1983 . A closer look at cloze :Validity and reliability . In J . W . Oller (ed .) . *Issues in Language Testing Research* ,pp . 237 -250 . Rowley :Newbury House .
- Carr ,T . H . & B . Levy . 1990 . *Reading and Its Development : Component Skills Approaches* . San Diego : Academic Press .
- Chihara ,T . ,J . W . Oller ,K . Weaver . & M . A . Ch vez -Oller . 1994 . Are cloze items sensitive to constraints across sentences ? *Language Learning* 27 /1 :63 -73 .

- Chavez-Oller, M. A., T. Chihara, K. A. Weaver, & J. W. Oller. 1994. When are cloze items sensitive to constraints across sentences? In J. W. Oller & J. Jonz (eds.), *Cloze and Coherence*, pp. 229-245. London: Associated University Press.
- Crehan, D. K., A. R. Koehler & J. S. Slakter. 1974. Longitudinal studies of test-wiseness. *Journal of Educational Measurement* 11/2: 209-212.
- Dechant, E. 1991. *Understanding and Teaching Reading: An Interactive Model*. Hillsdale: Lawrence Erlbaum.
- Ericsson, K. A. & H. A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press.
- Faerch, C. & G. Kasper. 1987. From product to process — Introspective methods in second language research. In C. Faerch & G. Kasper (eds.), *Introspection in Second Language Research*, pp. 5-23. Clevedon: Multilingual Matters Ltd.
- Feldmann, U. & B. Stemmer. 1987. Think aloud a retrospective data in Cloze taking: differ languages — different learners — same approaches? In C. Faerch & G. Kasper (eds.), *Introspection in Second Language Research*, pp. 252-253. Clevedon: Multilingual Matters Ltd.
- Green, A. 1998. *Verbal Protocol Analysis in Language Testing Research: A Handbook*. Cambridge: CUP.
- Markman, P. L. 1985. Rational deletion cloze and global comprehension in German. *Language Learning* 35/3: 423-430.
- Nunan, D. 1992. *Research Methods in Language Learning*. Cambridge: CUP.
- Shanahan, T., M. L. Kamil & A. W. Tobin. 1982. Cloze as a measure of intersentential comprehension. *Reading Research Quarterly* 17/2: 229-255.
- Shohamy, E. 1983. Interrater and intrarater reliability of the oral interview and concurrent validity with cloze procedure in Hebrew. In J. W. Oller (ed.), *Issues in Language Testing Research*, pp. 299-336. Rowley: Newbury House.
- Storey, P. 1997. Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing* 14/2: 214-231.
- Yamashita, J. 2003. Processes of taking a gap filling test: Comparison of skilled and less skilled EFL readers. *Language Testing* 20/3: 267-293.
- 陈晓扣、李绍山 2006 TEM 4 完型填空测试结构效度研究——答题过程分析法《现代外语》第1期。

(...continued from p. 28)

- Milanovic, M., N. Saville & S. Shen. 1996. A study of the decision-making behavior of composition markers. In M. Milanovic & N. Saville (eds.), *Performance Testing, Cognition and Assessment: Selected Papers from the 15th Language Testing Research Colloquium (LTRC)*, Cambridge and Arnhem, pp. 92-114. Cambridge: Cambridge University Press.
- Nakamura, Y. 1996. A study of raters' scoring tendency of speaking ability through verbal report methods and questionnaire analysis. *Journal of Communication Studies* 5: 3-17.
- Orr, M. 2002. The FCE speaking test: Using rater reports to help interpret test scores. *System* 30: 143-154.
- Rubin, H. J. & I. S. Rubin. 1995. *Qualitative Interviewing: The Art of Hearing Data*. Thousand Oaks, London, New Delhi: Sage.
- Sakya, A. A. 2000. Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. In A. J. Kunnan (ed.), *Fairness and Validation in Language Assessment*, pp. 129-152. Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.
- Shohamy, E. 1983. The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning* 33/4: 527-540.
- Shohamy, E. 1988. A proposed framework for testing the oral language of second foreign language learners. *Studies in Second Language Acquisition* 10: 165-179.
- Vaughan, C. 1991. Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (ed.), *Assessing Second Language Writing in Academic Contexts*, pp. 111-126. Norwood, New Jersey: Ablex Publishing Corp.
- Weigle, S. C. 1994. Effects of training on raters of ESL compositions. *Language Testing* 11: 197-223.