

Chapter 4

Results and Discussion

The following chapter presents a thorough elaboration of results and discussion in response to the listed research questions. Section 4.1 addresses the relationship between note contents and listening comprehension performance. Following section 4.1, section 4.2 further discusses which criterion can serve as the best predictor of EFL learners' listening performance and section 4.3 examines how question types interact with the subjects' test performance. Section 4.4 then presents the results of the influence exerted by the chance of reviewing notes. The recall performance of the non-reviewing group is reported in section 4.5 along with a summary of Chapter 4 in section 4.6.

4.1 What kind of notes can contribute to a better comprehension?

With regard to the post-listening comprehension questions as shown in Table 4-1, the subjects on average answered 10.90 questions correctly out of the 15 questions with the standard deviation of 2.00. The maximum of correct responses obtained by the subjects (NRV9, NRV19, RV1, RV17) was 14 while the minimum was as low as 7 (RV11).

Table 4-1: Subjects' Performance on the LC Test

Subjects' Correct Responses			
Mean	SD	Maximum	Minimum
10.90	2.00	14	7

Note: The mean score ranged from 0 to 15.

As far as the subjects' note contents are concerned, Table 4-2 lists the ranges of the six major criteria investigated in the present study coupled with their short forms for later use. TW stands for total of the words, IU for information units, TA for test-answerability, EF for efficiency, CP for completeness, and MI for major information.

Table 4-2: The Ranges of All the Criteria

Category		Range
1	total of the words (TW)	0-1884
2	information units (IU)	0-119
3	test-answerability (TA)	0-10
4	efficiency (EF)	0-1
5	completeness (CP)	0-1
6	major information (MI)	0-27

Table 4-3 further presents the subjects' means of each criterion, and their standard deviations. In addition to the above, the correlation results in interaction with the LC test are listed.

Table 4-3: The Means of the Six Criteria of the Subjects' Note Contents

Category	Mean	SD	<i>r</i>	<i>p-value</i>
TW	182.00	40.00	0.24	0.14
IU	46.00	11.00	0.38*	0.01 (p<0.05)
TA	6.50	2.00	0.59*	0.00 (p<0.01)
EF	0.25	0.04	0.24	0.12
CP	0.38	0.09	0.38*	0.01 (p<0.05)
MI	17.00	4.60	0.49*	0.00 (p<0.01)

Note: '' stands for significance.*

In terms of TW, that is, total of the words, the number of the words noted down on the subjects' note-taking sheet averaged out 182 words, ranging from the maximum of 272 words (RV8) to the minimum of 111 words (RV11), and its deviation standard turned out to be 40. The mean of information units, IU, recorded by the subjects was 46 information units with a top end of 77 (NRV10) and a bottom end of 20 (RV11). With respects to the third criterion, test-answerability, 2 subjects (RV6, RV8) achieved a full mark of 10 while another 2 subjects (NRV20, RV11) noted down as few as 2 information units that could assist in answering comprehension questions. Averagely, the subjects wrote down 6.5 information units that facilitated their answering comprehension questions. The fourth criteria, efficiency, which was a result of dividing IU by TW, yielded the mean, 0.25 with a standard deviation of 0.04, indicating that almost all the subjects did not differ greatly from each other. The statistical results ranged from 0.36 (RV6) to 0.18 (RV10). The completeness of the subjects' notes reached 0.38 and its standard deviation was 0.09. Among all the subjects, the notes could be as complete as 0.64 (NRV9) while the minimum was 0.16 (RV11). Concerning MI, major information units with a total of 27

averaged out 17 and the standard deviation was 4.6. NRV9 recorded 24 major information units, which was the greatest number, while RV11 wrote the fewest number of major information units, 4.

The raw data were then followed by an insight into how each criterion interacted with the subjects' LC test performance. The results in the category of each criterion were considered in relation to the subjects' LC test performances and their correlations were reported statistically with their regression results and p-values.

Criterion 1: Total of the Words

The correlation between the first criterion, the total of the words and the subjects' listening comprehension was found by the regression, 0.236. The statistic result apparently indicated that these two factors were insignificantly relevant. In other words, more of the words written down by the subjects might not directly lead to a better comprehension. This finding could be supported by examining the subjects who answered more than 90% of the questions correctly. The total of the words noted down by them were 244 (NRS9), 173 (RVS1), 226 (NRS19), 175 (RVS17), respectively, showing a wide range, as can be seen in Table 4-4.

Table 4-4: TW Scores of the Subjects with Over 90% of Correct Responses

Subjects	NRS9	RVS1	NRS19	RVS17
TW score	244	173	226	175

Another insight into the subjects whose correct responses were 60% could be another piece of supportive evidence, as shown in Table 4-5.

Table 4-5: TW Scores of the Subjects with Less than 60% of Correct Responses

Subjects	RVS7	NRS20	NRS6	RVS18	NRS7	NRS11	NRS17
TW score	251	214	189	176	173	157	136

These subjects scribbled 251 (RVS7), 214 (NRS20), 189 (NRS6), 176 (RVS18), 173 (NRS7), 157 (NRS11), and 136 (NRS17) words respectively, indicating that it was almost the same distribution as the group previously discussed. The same number of words failed to indicate the subjects' listening performance. From the statistic result, a correlation between TW and listening comprehension performance lacked a strong connection to conclude their relationship. However, such results were contradictory to the findings of the pilot study in which TW indeed showed a significant correlation with the subjects' listening comprehension. The reason could be attributed to the proficiency of the subjects. In the pilot study, only eight out of forty-one subjects were examined and they represented two extreme groups whose English proficiency differed considerably. However, in the present study, the subjects' English proficiency ranged widely from the high to low levels.

The finding of the present study countered Nye's (1978) claim that between the total of the words and test scores existed a small yet significant positive correlation. In addition, the finding failed to support Norton (1981) who stated that the total of words significantly index the subjects' test performance. Though the present study did not detect a significant positive correlation, a negative correlation was not identified, either. Dunkel (1988) contended that too much writing, which lacked efficiency, was inversely correlated with the subjects' performance. Such a contention was not found true in the present study. What is found corresponds to Chaudron, Loschky, & Cook's (1994) conclusion that the total of words could not significantly index the subjects' test performance.

All the interviewees reached the consensus that the number of recorded words lacked a solid connection with their test performance. As NRV9 said, recording more words was just an act of wasting time. Moreover, it prevented the listeners from receiving upcoming messages as RV7 reported.

Excerpt 4-1¹

1. I don't think it wise to try to write down many words. First of all, *doing so simply means wasting time*. Second, it doesn't help me organize the ideas in a logical way. (taken from NRV9)
2. Verbatim recording doesn't mean getting the key information. That is, when I write down every single word given by the lecturer, I may not get the information that is helpful in the later test. On the other hand, *the time limit doesn't allow me to write so many words*. (taken from RV7)
3. Writing too much only *distracts me from receiving the information conveyed by the speaker*. If I can *comprehend a certain piece of information, I will be able to use my own words to record it concisely*. (taken from RV3)

From the perspective of the interviewees, the verbatim recording causes inefficiency and distraction, which might prevent the subjects from receiving the upcoming information. Another reason for not doing word-by-word recording could be attributed to time limit. Hence, writing down as many words as possible was not considered as a wise note-taking skill by the interviewees.

Criterion 2: Information Unit

IU, information unit, was found to be significantly correlated with the subjects' test performance with a regression of 0.381 and a p-value of 0.014. This could be proved to represent a significantly positive correlation. Generally speaking, those who

¹ All the interview data were in Chinese and translated to English by the researcher.

recorded more than the average information units outperformed those who noted fewer than the average, as in Table 4-6.

Table 4-6: LC Test Score of the Subjects

Subjects	Above the IU average (n=18)	Below the IU average (n=23)
LC test score	11.61	10.34

To some extent, more information units could predict the subjects' better test performance while fewer information units could serve as a precursor to the subjects' worse test performance.

Investigating the performances of the two extremes separately enabled us to notice a wide gap between their number of information units. The subjects whose percentage of correct responses exceeded 80% on average noted down 50.40 information units while the low achievers who answered fewer than 60% of the questions correctly only contained 36.60 information units, as shown in Table 4-7. This difference supported our significant correlation.

Table 4-7: IU Scores of the Subjects

Subjects' correct response	Above 80% (n=12)	Below 60% (n=5)
IU score	50.40	36.60

The positive correlation added more evidence to Palkovitz & Lore's (1980) and Baker & Lombardi's (1985) studies. It also corresponded to Harris's (1974) findings

that more information units found in the subjects' notes ensured better test scores. However, though many researchers argued for such a correlation, Dunkel (1988) and Chaudron, Loschky, & Cook (1994) found that the subjects' test performance was in short of a significant correlation with their number of information units. The present study apparently argued against their results with a positive correlation.

The correlation between information units and test performance was further testified in the interviews. Of all the four subjects, only one (i.e. RV11) considered information the most important criterion. The subject reassured herself with the seizure of more information in the test, as can be seen in the following excerpt:

Excerpt 4-2

For me, I think more information can guarantee me a good test performance. When I am able to cover as much information as possible, *I feel reassured and confident* in the comprehension tests. (taken from RV11)

. However, the rest of the interviewees simply reported that information units only deserved some importance and not to a decisive extent. Therefore, both the interview data and statistical results authenticated the significant yet minor importance of IU.

Criterion 3: Test-answerability

TA, test-answerability, turned out to be the most correlated criterion to the subjects' test performance. The regression score, 0.592, was the highest among all the criteria. This high relevancy indicated that the subjects with more answerable information units were more likely to achieve a better performance. On the contrary, those with fewer answerable information units tended to get a lower score. This contention can be reinforced with a closer look at the bottom five subjects' notes in

which there were only 2 (RV11), 3 (NRV13), 5 (RV13), 5 (RV20) and 6 (RV2) test-answerable information units found in their notes, as in Table 4-8. It is no surprising that they scored lower than the high achievers who recorded 7 to 9 answerable information units on average. Therefore, test-answerability was indeed the most relevant criterion to the subjects' test performance.

Table 4-8: TW Scores of the Bottom 5 and Top 5 Subjects

Subject	Bottom 5					Top 5				
	RV11	NRV13	RV13	RV20	RV2	NRV9	RV1	NRV19	RV17	NRV2
TW score	2	3	5	5	6	9	7	7	9	9

The present study further solidifies the contention that test-answerability, though seeming eligible, appears to be the best indicator of the subjects test performance (Dunkel, 1988). Compared with Dunkel's (1988) study, which designed a different task of lecture concept recognition, it somehow can be considered that test-answerability indeed is quite an indicator of the subjects' test performance regardless of the tasks such as multiple choices or recognition task.

All the interviewees agreed that the criterion of test-answerability determined their test performance. As NRV9 pointed out, the answerable information was the most important reference for them when they took a LC test.

Excerpt 4-3

1. When I can refer to my notes and *find out more test-answerable information, I surely will get a better score*. If possible, I will try to go over the questions and take notes of the needed information. (taken from NRV9)
2. Due to the training in high school for “*test-wiseness*,” *I always take a glance over all the questions* and try to listen for “possible” information to note down. (taken from RV3)

Obtaining needed information in the notes was considered an important criterion of determining their good test performance. For enhancing the test-answerability, the interviewees pointed out the positive effects of previewing listening comprehension questions.

Criterion 4: Efficiency

EF, efficiency, failed to establish a correlation with the subjects’ test performance. The regression result turned out to be 0.244 coupled with a p-value of 0.124. As a matter of fact, the standard deviation for all the subjects appears was low to exert any influence, showing that homogeneity was found among the subjects. The statistical analysis reinforced the claim with the regression result, 0.244, which stood for no significant relevancy. As a result, efficiency was not one of the factors that could decide the subjects’ test performance.

The present findings appeared to be different from Howe’s (1970) statement that efficiency, which embeds the most information units by jotting down the fewest words, is a factor in good notes. Though this statement was supported by Aiken, Thomas, and Shennum’s (1975) study, such a connection between efficiency and the subjects’ test performance was not found in the present study. This finding is

consistent with Dunkel (1988) and Chaudron, Loschky, & Cook (1994), showing that no significant positive correlation between the criterion of efficiency and the subjects' test performance.

Efficiency, according to the interviewees, appeared less prominent in affecting their test performance. The interviewee placed less emphasis on being efficient than receiving correct information, as can be seen in Excerpt 4-4.

Excerpt 4-4

Being efficient doesn't mean being correct. For another thing, if I can catch the major line of the information, there appears no need to be efficient. (taken from RV11)

Criterion 5: Completeness

CP, completeness, was significantly yet slightly correlated with the subjects' test performance. Among all the subjects, the average percentage of the completeness was 0.038 with a standard deviation of 0.1. This result manifested a slight yet significant variation. The subjects who answered more than 80% of the questions correctly reached 0.42, while those whose correct responses were less than 60% achieved only 0.31, as in Table 4-9. This difference again proved that more completeness of notes could significantly exert its positive effect on the subjects' test performance.

Table 4-9: CP Scores of the Subjects

Subjects' correct response	Above 80% (n=12)	Below 60% (n=5)
CP score	0.42	0.31

The present findings shared the same results as what Locke (1977) identified as the positive relationship between his students' note completeness and their final course grades. On this ground, note completeness indeed to some extent could indicate the subjects' performance. Nonetheless, our findings reversed the view held by Dunkel (1988) and Chaudron, Loschky, & Cook (1994) that there was no significant correlation between completeness and test performance.

Though completeness was slightly significant, one of the interviewees did not credit her good test performance to this criterion.

Excerpt 4-5

More complete information may *fail to represent all the necessary information for comprehension*. So I think I don't have to try to reach completeness of notes but associate what I have heard with my background knowledge. (taken from NRV9)

NRV9 thought that the effort to reach completeness might result in the inability to comprehend the whole lecture and result in some missing information.

Criterion 6: Major Information

MI, major information, was more relevant to the subjects' test performance than IU, information unit, statistically. The regression result turned out to be 0.49, showing some relevancy. The subjects who got more than 80% of the correct responses jotted down 20.08 units averagely out of 27 major information units while those who achieved less than 60% correct responses only had 13.60 pieces of major information in their notes. The latter group almost lacked a quarter of all the major information units, suggesting the former group could catch more information units than the latter group. This difference was also reflected in test performance of the two groups. With more major information, the above 80% group answered 13.33 questions correctly. In

contrast, the below 60% group only achieved 7.8 correct responses, as shown in Table 4-10.

Table 4-10: MI and LC Scores of the Subjects

Subjects' correct response	Above 80% (n=12)	Below 60% (n=5)
MI score	20.08	13.60
LC score	13.33	7.80

Major information has been found to be significantly correlated to the subjects' information units. This implicitly showed that the high achievers were more aware of the structure of a lecture and were more capable of catching the main ideas. Unlike the high achievers, the low achievers seemed to rely more on bottom-up strategies to comprehend the lecture. Such findings contradicted Connor's (1984) claim that it was "minor points" not "main points" that was positively correlated to test performance. Apparently, the criteria of major information units indicated the subjects' test performance better than information units. As a result, the ability to identify the major information appeared to be of more value than the act of writing down any information.

An agreement that major information decisively affected the subjects' test performances was reached by the interviewees. As RV7 mentioned, writing down major information was efficient in gaining a general picture of the whole lecture. RV11 even regarded major information as an association with background knowledge to supplement missing gaps in comprehension of the lecture.

Excerpt 4-6

1. Within a limited amount of time, jotting down major information can help me best in doing the comprehension questions. When the input continues to flow in, *it's impossible for me to write down everything*. Then the only solution to frame the main picture of the lecture is major information. (taken from RV7)
2. I think taking notes of major information means comprehending key information of the lecture. And on the basis of major information, I will be able to *associate it with supplementary information* and then to get the whole picture. (taken from RV11)

Writing major information was considered workable and feasible when an abundance of information flowed in. On the other hand, the subjects' major information could serve as their frame of the whole lecture and the association between major information and background knowledge then could flesh out the frame.

Examining the six criteria, it was concluded that IU, TA, CP, and MI were significantly correlated with the subjects' performance. Such results differed from Dunkel's (1988) findings that only test-answerability exerted its significant effect on test performance. The difference could be attributed to the material Dunkel adopted. It was a lecture that lasted around 22 minutes. This length outnumbered the lecture used in the present study by the margin of 10 minutes. Faced with a lecture of such length, the subjects then were supposed to write down a large number of information units, which might cause them to feel exhausted and lose focus. On the other hand, when answering the same number of questions – 15, the subjects had more difficulty identifying and locating information needed for the test questions. In contrast, test-answerability was found to be the best predictor since it directed the subjects to

the necessary clues in answering the comprehension questions. Unlike Dunkel's (1988) lecture length, the present study adopted a 12-minute lecture, which less exhausted the subjects and exposed them to a smaller quantity of information units. Therefore, IU and CP were significantly correlated with the subjects' test performance.

After examining the six criteria, it was found that IU, TA, CP, and MI were significantly correlated with the subjects' performance. In essence, these four criteria shared some properties. For example, more information naturally ensured more completeness, and more major information which could be subcategorized into information units surely guaranteed more information units. As to test-answerability, the goal of a test commonly focuses on major information. This is why it was correlated with major information in the present study.

4.2 Among all the criteria of note contents, which ones are the best predictors of EFL learners' listening performance?

The previous discussion has initially probed the relationship between the subjects' notes and their listening performance. In order to gain a more comprehensive picture, this section intends to direct the attention to the difference between the high achievers and the low achievers of the present study. In this section, top 20% of the subjects were considered the high achievers and bottom 20% were the low achievers. The high achievers consisted of 12 subjects, who answered more than 13 out of 15 questions correctly. The low achievers were 12 subjects who answered no more than 9 questions correctly.

By retrieving the data from these two groups, the relationship between each of the criteria and the subjects' listening performance was examined. Table 4-11 showed the average LC test score and note contents of the high achievers in comparison with that of the low achievers'.

Table 4-11: Subjects' LC Test Scores and Note Contents

Group		High Achievers		Low Achievers	
Category		(n=12)		(n=12)	
LC test scores		Mean	SD	Mean	SD
		13.33	0.49	8.50	0.67
Note contents	TW	190.00	41.60	171.00	36.30
	IU	50.40	8.54	38.10	9.33
	TA	8.00	0.85	4.75	1.76
	EF	0.27	0.03	0.23	0.05
	CP	0.42	0.07	0.32	0.08
	MI	20.08	2.68	14.08	5.33

Table 4-12 indicated the results of the correlation between the high achievers' and low achievers' performance and each criterion.

Table: 4-12: Correlation Results of the Two Achieving Groups and the Whole Group

Criteria	Correlation			
	high achievers and low achievers		all the subjects	
	<i>r</i>	<i>p-value</i>	<i>r</i>	<i>p-value</i>
TW	0.341	0.103	0.236	0.137
IU	0.631*	0.001 (p<0.01)	0.381*	0.014 (p<0.05)
TA	0.798*	0.000 (p<0.01)	0.592*	0.000 (p<0.01)
EF	0.453*	0.026 (p<0.05)	0.244	0.124
CP	0.631*	0.001 (p<0.01)	0.381*	0.014 (p<0.05)
MI	0.618*	0.001 (p<0.01)	0.492*	0.001 (p<0.01)

Note: '*' stands for significance.

Compared with the results of the entire group of the subjects, the results presented even more evident and stronger correlation between the subjects' listening comprehension performance and each of the criteria on the whole. After removing the performance of the subjects who performed averagely, all of the criteria were in uniformity with our previous findings but achieved an even stronger relation with the listening performance. Among all the criteria, TA, test-answerability, appeared highly correlated with the subjects' performance. The high achievers recorded averagely 8 clues that reveal the answers to the local questions. In contrast, in the low achievers' notes, there were only 4.75 clues that could lead them to choose the correct answers, as can be seen in Table 4-11. The result reinforced Dunkel's (1988) claim that only the test-answerability could predict the subjects' listening performance. In other words, in our subjects' notes, the more test-answerable information was identified, the better they performed.

Second to TA, test-answerability, was IU, information unit. The result almost turned out to be twice as relevant to the subjects' performance as the regression result from all the subjects. The criterion of information units could serve as a more precise indicator of the low achievers' and high achievers' performance than that of the entire group. The result almost reached a high level of correlation. The high achievers scribbled 50.40 information units on average in their notes, but the mean number of the information units for the low achievers was only 38.1. Therefore, it can be concluded that the high achievers succeeded in receiving input and transforming them into their notes while the low achievers with the difficulty comprehending the given messages failed to note down more information units. The present findings reinforced the correlation between the subjects' information units in their notes and their test performance, which echoed the findings of Palkovitz & Lore (1980) and Baker & Lombardi (1985) as well.

Major information was proved to be more correlated to the subjects' performance after depriving the intermediate subjects' performance of the analysis. Similar to IU (information unit) major information was highly relevant to the subjects' performance. The mean number of major information units obtained by the high achievers was 20.08 while in the low achievers' notes, there were only 14.08 pieces of major information, as in Table 4-11. A greater number of major information can be a predictor of the subjects' performance. However, the identified strong correlation between major information and test performance contradicted Connor's (1994) contentment that minor points rather than main points were positively correlated to test performance. According to our findings, major information and the LC test were closely related.

Different from the previous findings, EF, the criterion of efficiency, was found crucial in the present analysis. Though significantly yet slightly correlated with the subjects' performance, efficiency indicated the subjects' performance to a certain extent. The high achievers reached the efficiency of 0.27, that is, every single recorded in their notes was loaded with 0.27 key piece of information. As far as the low achievers were concerned, their average efficiency was 0.23. Such a contrast showed that the high achievers were mostly efficient in taking notes. High efficiency identified in the high achievers added support to Howe's (1970) findings that more efficient subjects were able to obtain better scores.

The only criterion that remained insignificantly correlated was TW (the total of the words). On the basis of the findings, the total of the words was the least useful predictor in the subjects' test performance. The more words the subjects wrote down on the note-taking sheet did not guarantee their good performance (Chaudron, Loschky, & Cook, 1994; Dunkel, 1988; Howe, 1970).

4.3 What is the relationship between note contents and global and local questions?

For the purpose of examining the subjects' understanding on the lecture, the comprehension questions in the LC test consisted of 5 global questions and 10 local questions. Table 4-13 showed the results of the subjects' performance on global and local questions. The subjects averagely answered 3.8 out of the 5 global questions and 7.09 out of the 10 local questions correctly and the percentage of correct responses to the global questions was 76% and that to the local questions was 70.9%.

Table 4-13: Subjects' Performance on Global and Local Questions

Question Type	global	Local
Range	0-5	0-10
Mean	3.80	7.09
Percentage of correct responses	76%	70.9%
Standard Deviation	0.70	1.60
(n=41)		

After checking the subjects' responses, on the part of the global questions, most of the subjects only failed to answer one global question correctly, but they failed to answer three local questions on average. By examining the standard deviation of the global questions, which was 0.7, the subjects did not differ from each other greatly. However, the subjects performed much more differently on the local questions with a standard deviation of 1.60. Therefore, it is evident that the type of global questions received a higher percentage of correct responses from the subjects and also caused little within-group variations.

Shohamy & Inbar (1991) stated that responses to local questions were more

readily attainable than global questions, which is countered by Chen's (2003) findings. In the present study, the percentage of correct responses to global questions appeared a little bit higher than that of local questions. However, after running the t-test, the p-value turned out to be 0.053, indicating that there was not any significance between the two types of questions, as can be seen in Table 4-14.

Table 4-14: T-test Results of Correct Response Percentage of Global and Local Questions

t	DF	<i>p-value</i>
-1.994	40	0.053

After depriving of the task factor asserted by Chen (2003), in the same format of multiple choices, global questions received almost the same difficulty level as local questions. Therefore, the question type did not correlate with difficulty level in a post-lecture listening comprehension.

Table 4-15 shows the correlations between each criterion and two question types – global and local:

Table 4-15: The Correlations between Each Criterion and Global questions

Question Types Criteria	Global Question		Local Question	
	<i>r</i>	<i>p-value</i>	<i>r</i>	<i>p-value</i>
TW	-0.19	0.91	0.31	0.05
IU	0.06	0.71	0.45*	0.00
TA	0.27	0.10	0.62*	0.00
EF	0.10	0.53	0.26	0.10
CP	0.06	0.71	0.45*	0.00
MI	0.20	0.20	0.53*	0.00

*Note: * stands for significance. $p < 0.01$*

Based on the results, none of the criteria was found to be able to index the subjects' performance on the global questions, implying that note-taking appeared to lack an interplay with global questions. Since global questions were mainly designed to test the subjects' general understanding without referring to any specific detailed information, it was not necessary for the subjects to record the information they received. Among all the criteria, though insignificantly, the most relevant criterion amazingly was test-answerability. This finding can be explained through a correlative analysis of global and local questions. The correlation between the performance on the global questions and the local questions reached a significant level with the regression, 0.41 with the p -value, 0.01, as shown in Table 4-16. This revealed that higher achievement on the local questions implied a better performance on the global questions.

Table 4-16: Correlation between Global Questions and Local Questions

Question Type	<i>r</i>	<i>p-value</i>
Global and Local Questions (n=41)	0.41*	0.01

Note: ‘’ stands for significance. $p < 0.01$*

Major information, though insignificant, was ranked as the second most relevant criterion and comparatively established a little bit relevance with the subjects’ test performance on global questions. This may be due to the fact that a larger amount of major information can bring forth a better general picture of the lecture.

On the part of local questions, of the six criteria under investigation, TA, test-answerability, was the most relevant criterion related to the subjects’ test performance on local questions. The regression result was 0.62, which almost hit a level of high correlation. Therefore, this finding indicated that when the subjects were able to record the information that was needed to answer the comprehension questions, they would have a greater chance to perform better. Through examining each individual performance of the high achievers, it was found that the subjects could answer more questions correctly than the number of recorded test-answerable information units. Generally speaking, the high achievers could answer more local questions correctly than what their notes could provide. As for the low achievers, on average, the number of their correct responses to local questions was closer to their test-answerability score. It seemed that test-answerability almost precisely predicted the number of the low achievers’ correct responses. The importance of test-answerability was also accentuated by most of all the interviewees. As NRV9 mentioned, the test-answerable information received the foremost importance and

previewing the questions could also guide her to take notes of test-answerable information.

Excerpt 4-7

When I have more test-answerable information, surely I will have a better achievement. Especially *if provided the chance of previewing the comprehension questions beforehand*, I will be more capable of recording the needed information and spare the effort of taking notes of redundant information. (taken from NRV9)

Previewing the questions in fact served as a direction for the subjects to take test-answerable notes out of a large quantity of information. By means of this, they were more likely to retrieve the necessary information to perform better on the listening comprehension test.

Major information was ranked as the second in relevance to the subjects' test performance on local questions. The correlation between major information and local questions also reached a significant difference. More major information guaranteed more correct responses to local questions. This result corresponded to the reflection of NRV9 that more major information could ensure her recollection of minor information through background knowledge. This also explains the poor use of major information and reasoning skills of the low achievers. RV7 stated in the excerpt below that major information was not her tool to activate the memory of minor information.

Excerpt 4-8

1. Recording major information serves to *be a guide to associate my background knowledge with some minor information*. Therefore, taking notes of major information, in my opinion, is quite important. (taken from NRV9)
2. I don't think taking notes of major information is useful because *those pieces of information usually are not the target comprehension questions in the test*. So major information seems to take no effect. (taken from RV7)

RV7's personal belief that test questions were usually related to detailed information led her to overlook the significance of major information. In addition, this belief contradicted most test-givers' belief that test questions should be aimed for major information. Such a wrong belief partly has contributed to her poor performance on the comprehension test.

In addition to the previous criteria, IU (information units) and CP (completeness) was significantly correlated with the subjects' test performance. The more completeness of the subjects' notes could provide a larger number of cues for the subjects to achieve correct responses. Completeness was regarded as the most important criterion when the interviewed low achievers were asked to rank all the criteria in order of the importance in relation to their test performance. According to RV7's, the limitation of memory load was the major concern which led her to remember the lecture through the recorded notes. When deprived of the chance of previewing the questions in advance, as RV7 said, she could only manage to record any possible information that might be related to comprehension questions. Therefore, the chance of previewing questions affected the subjects' test strategies.

Excerpt 4-9

1. When answering the questions aimed at specific information, I think completeness is indeed the most important. *Usually my memory can't be loaded with too much information without taking notes.* However, the local questions usually want to elicit detailed information. (taken from RV7)
2. *Without getting to know what will be tested in advance,* I think the more complete the information is, the better I can do. (taken from RV7)

As to the rest of the criteria, it was not surprising to find TW (the total of the words) and EF (efficiency) still lacked any correlation with the subjects' test performance. For the criterion of efficiency, the lack could be explained in that the subjects did not differ considerably in this respect. For the total of the words, RV3 claimed that more words were not accompanied with more needed information.

Excerpt 4-10

Recording more words only means a mess to me. *Without organization,* more words don't offer me a clear picture of what the lecture is about and even can confuse me. (taken from RV3)

Based upon the above excerpt, it was obvious that note-taking was correlated more with local questions than with global questions. All of the interviewees admitted the necessity of note-taking when answering local questions. All of them attributed this need to the limitation of their short-term memory.

The high achievers and the low achievers held different viewpoints on the necessity of note-taking on the part of global questions and local questions. Amazingly, the high achievers were in uniformity with what was found in the statistical analysis. That is, they credited the act of note-taking to only local questions

but not to global questions. However, the low achievers persisted in the need for taking notes on these two question types, even in answering global questions, as can be seen in Excerpt 4-11:

Excerpt 4-11

1. I don't think it necessary to take notes because the mental comprehension can help me answer the questions for general information. On the other hand, *these questions are not aimed at specific information*. I don't really need to take notes of any details. (taken from NRV9)
2. Basically, I think note-taking is not the major factor in answering the global questions but the topic. If *the topic is familiar to me, there stands no reason for me to take notes*. (taken from NRV9)
3. For general information, after listening to the lecture, I think *my memory can afford to work out a general picture about the main idea of the lecture*. In addition, I think the purpose of taking notes is simply to record details that will cause difficulty in recalling. Therefore, there is no need to take notes. (taken from RV3)

NRV9 reported that taking notes was the act oriented toward answering local questions. For general comprehension questions, the act then lost its ground since her mental comprehension could handle them, with which RV3 agreed as well. On the other hand, NRV9 further pointed out that faced with a familiar topic, she could simply apply her background knowledge to answer the questions without taking notes. Hence, NRV9 and RV3 both thought that the act of note-taking would be more useful in answering local questions.

For the high achievers among the interviewees, they both agreed that mental comprehension has provided enough clues to global questions and note-taking appeared to be redundant. However, the low achievers (RV7, RV11) and the high

achievers had an opposite view. As RV7 mentioned, note-taking was a habit deeply rooted in her, which compelled her to take notes even in response to global questions. On the other hand, both RV7 and RV11 reached the consensus that the building of the parts could contribute to a whole, which was a bottom-up strategy.

Excerpt 4-12

1. I think it necessary to take notes. For one thing, *it has been a deeply-rooted habit to take notes when I am exposed to any listening material.* (taken from RV7)
2. For another, more complete notes can contribute to a better comprehension, which in turn *brings forth positive effects on answering global questions.* (taken from RV7)
3. For me, taking notes is indispensable. I often miss the whole picture and fail to catch the gist of the lecture. So *taking notes can help me build the main frame of the lecture in my mind.* This way, I will be able to comprehend the main idea of the lecture and have more clues to the global questions. (taken from RV11)

The different opinions between the high achievers and low achievers seemed to indicate their different comprehension processing. The high achievers seemed to adopt the top-down processing. In contrast, the low achievers seemingly could not handle the top-down processing and resorted to the bottom-up processing. Such variations resulted in their necessity of taking notes.

In the previous discussion, global questions were found to lack any correlation with all the criteria. However, after removing the performance of the intermediate subjects, the results shown in Table 4-17 diverged from the previous findings. IU, TA, CP, and MI all established a significant yet slight correlation with the high and low achievers' performance on the global questions.

Table 4-17: Global and Local Questions Correlation of the Two Achieving Groups

Criteria	High Achievers and Low Achievers (n=24)			
	Global Questions		Local Questions	
	<i>r</i>	<i>p-value</i>	<i>r</i>	<i>p-value</i>
TW	0.19	0.38	0.36	0.09
IU	0.41*	0.05 (p<0.05)	0.64*	0.00 (p<0.01)
TA	0.59*	0.00 (p<0.01)	0.78*	0.00 (p<0.01)
EF	0.30	0.15	0.45*	0.03 (p<0.05)
CP	0.41*	0.05 (p<0.05)	0.64*	0.00 (p<0.01)
MI	0.45*	0.03 (p<0.05)	0.61*	0.00 (p<0.01)

Note: ‘’ stands for significance*

Of the 4 relevant criteria, test-answerability came in first though it actually ran short of a direct relationship with global questions. As previously discussed, the test-answerability best predicted the subjects’ performance on local questions. With a high correlation between global and local questions, test-answerability indirectly functioned best as a predictor of the subjects’ performance on global questions. Major information and information units shared a similar correlation with their test performance. These two criteria indicated a better comprehension and a better construction of the whole picture of the lecture. Apart from the above four criteria, TW and EF were found unrelated to the subjects’ test performance although their regression results increased in comparison with the original results. Therefore, in consistency with the previous discussion, the total of the words and efficiency failed

to predict the subjects' performance on global questions.

With regard to local questions, without considering the performance of the intermediate subjects, the results showed an even stronger correlation between five out of the six criteria and the performance of the high and low achievers on local questions (see Table 4-17).

In comparison with the results from the entire group, all the correlation results were revised upward to quite an extent. In order of relevancy, TA, IU, CP, MI, and EF were correlated to the subjects' test performance on local questions, which did not differ greatly from the previous findings. Test-answerability was highly correlated with the subjects' performance, which meant it could serve as the most precise indicator of the subjects' performance. IU, information units, was found to be highly correlated with the subjects' test performance. The high achievers were able to keep track of the given information more than the low achievers, resulting in their greater performance on local questions. Like IU, MI, major information was also highly correlated with the subjects' performance. EF amazingly had a slight correlation to the test performance. The exclusive criterion that lacked a significant correlation went to TW. The total of the words remain the worst predictor.

With regard to the global questions, the findings of the present study did not agree with Nye's (1978) claim that the number of words was positively correlated with the subjects' test score. The same was true of local questions. Neither responses to global questions nor local questions could support Norton (1978), who stated that the total of words could predict the subjects' performance. The insignificant relationship between global or local questions and the subjects' test performance echoed Dunkel (1988) and Chaudron, Loschky, & Cook (1994), who held the view that the total of the words was not the factor that would decide the subjects' performance.

The criterion of information unit has established a correlation with the subjects' performance on local questions but not on global questions. The significant correlation between local questions and information units consisted with Harris's (1974) statement that an increasing number of information units resulted in the subjects' better recognition score. The present findings could further support Connor's (1984) contentment that it was the minor points not main points that were positively related to the subjects' test performance. This explained why the subjects relied more on the notes in responding to local questions.

Dunkel (1988) suggests that test-answerability was proved to be the most relevant criterion to his subjects' test performance. Based upon our findings, test-answerability indeed was closely related to our subjects' performance on local questions. As to global questions, the irrelevance contradicted Dunkel (1988) but affirmed Chaudron, Loschky, & Cook (1994).

The fourth criteria, efficiency failed to correlate itself with either global or local questions. Although Howe (1970) stated that the efficient note-taking ensured a better performance, the present study does not support this statement but was consistent with Dunkel's (1988) claim that there was no significant correlation between efficiency and test performance. With regard to the criterion of completeness, our subjects' performance on local questions had a positive correlation like what Locke (1977) found. However, Locke's (1977) findings were not totally supported in the case of global questions.

In general, global questions did not appear correlated with any criteria but after removing the intermediate students, the criteria that were correlated were still TA, MI, IU, and CP like our results found in the LC test. As to local questions, in addition to the previous four criteria, EF appeared slightly correlated but TW failed to associate itself with the subjects' test performance on local questions.

4.4 Can the subjects benefit from note reviewing

Table 4-18 illustrates the results of the delayed recall task. The non-reviewing group and reviewing group, whose listening performances were approximately the same, differed considerably from each other. The non-reviewing group, deprived of the chance of reviewing the notes one week after listening to the lecture, on average recalled only 8.61 information units in their recall task (see Table 4-18).

Table 4-18: Subjects' Performance on the Two Tasks

Group \ Score	The LC task		The Recall task	
	Mean	SD	Mean	SD
Non-reviewing (n=21)	11.00	1.87	8.61	3.83
Reviewing (n=20)	10.80	2.17	16.55	6.75

Only two (NRV2, NRV9) out of the 21 non-reviewing subjects succeeded in recollecting more information units than the reviewing group and these two subjects happened to be the best performers on the LC test. In contrast to the non-reviewing group, the chance of reviewing notes has contributed to a double number of the information units reconstructed by the reviewing group. On average, the reviewing group showed 16.55 information units in comparison with the non-reviewing group. Under the condition of similar listening proficiency, the chance of reviewing notes one week after the lecture significantly affected the subjects' performance.

The significant difference exerted by the chance of reviewing notes found in the interviews with the four subjects as well. All the interviewees agreed that the

provision of notes would facilitate their reconstruction of the lecture. As NRV9 said, the limitation of memory should be compensated for by the reviewing of notes, and otherwise, the uncertainty of lecture construction would in turn arise.

Excerpt 4-13

I think the *provision of notes* indeed can assure me a better reconstruction of the lecture because *I don't think my memory is so reliable after one week*. Referring to the notes can prove whether my recall of any specific information is correct or wrong. If deprived of the chance of reviewing notes, the only resort that remains is my memory and *suspicion of my reliability of my memory* arises in turn. (taken from NRV9)

RV7 reasserted the assurance that her notes could help complete the follow-up task to summarize the lecture.

Excerpt 4-14

When I have the chance to review notes, I can recall what I recorded and surely can recall more information than when I don't have the chance. Without the chance to review, I can only try hard to recall some *general information* and will *feel insecure about what I have recalled* since I can't make a clear-cut judgment of whether it is right or wrong. (taken from RV7)

RV7 even reported how the opportunity of reviewing notes affected her recall performance, as can be seen in the following excerpt:

Excerpt 4-15

When I am given the chance of reviewing notes, I think around 70% to 80% of the information recorded in my notes can be reconstructed successfully. But if I am forbidden to refer to my notes, then *what remains in my mind would be some very general and vague impressions about the lecture*. For the details, I may need notes to recall. (taken from RV3)

RV11 indicated the facilitating effect of referring to the notes in the interview, as shown in Excerpt 4-16:

Excerpt 4-16

1. When I can refer to my own notes, I think I will be able to reorganize my recorded information in a sequential order and even *supplement the original information with more additional unrecorded information* to make the whole construction of the lecture more complete. (taken from RV3)
2. However, if deprived of the chance of reviewing notes, I may go blank in my brain and have to search my memory for possible information. *I may not be able to organize my recalled information in a logical and sequential order*. What can be recalled probably is some information that I think is interesting. (taken from RV3)

The four excerpts from the interviews indicated that the chance of reviewing notes could assist in (1) recalling detailed information, (2) organizing the ideas in a sequential order, (3) making sure the recalled ideas are correct, and (4) making up for the limitations of short-term memory.

On the basis of the results coupled with the interviews, the function of

note-taking was highlighted. As literature has identified the major two functions of note-taking – encoding and external storage (DiVesta & Gray, 1972), the present study added more support to the second function – external storage, which serves to associate input with background knowledge. It was found that with notes serving as a resource for later review, the subjects could generate more information that had previously been stored in their memory. This explained why the subjects with notes to review outperformed those without. The facilitative effect of note-taking on test performance identified in the present study follows the general agreement that the chance of review indeed leads to a higher achievement on the subjects' test performance (Hartley, 1983; Kiewra, 1985).

The effects of note-taking on recall are generally explored in two directions – immediate recall and delayed recall. On the part of immediate recall, DiVesta argued that the opportunity of reviewing notes along with the action of note-taking enhanced the subjects' test performance on the immediate recall task. However, the function of external storage has been challenged and questioned due to the effect of short-term memory (Dunkel, Mishra, & Berliner, 1989). Dunkel, Mishra, & Berliner (1989) stated that their subjects relied more on short-term memory rather than the action of note-taking in recalling immediately after the lecture. In Chaudron, Loschky, & Cook's (1988) study, the immediate recall failed to demonstrate any effect of note retention. The note-retention group and non-note-retention group did not exhibit any difference. Therefore, the exploration of immediate recall has been transformed into that of delayed recall. The present study was proved to strengthen Dunkel, Mishra, & Berliner's (1989) statement that a chance to review notes to take notes serves a key factor to augment learning. By excluding the effect of short-term memory, the present study adopted the delayed recall which was conducted one week after the lecture. The subjects' performance on the delayed recall was significantly affected by their chance

of reviewing notes. The reviewing group averagely recalled double as many information units as the non-reviewing group. However, the present study lacked solid evidence to support Dunkel, Mishra, & Berliner's (1989) argument for the chance to review notes rather than the chance to record notes that takes effect. Based upon the previous discussion, our subjects' note contents were proved to demonstrate their effect on their listening comprehension performance. The test performance was highly correlated with their note-taking. Therefore, in the present study, both encoding and external storage were found to interplay to take effects instead of excluding one from the other.

4.5 What kind of notes can contribute to a better retention?

In order to examine the subjects' retention, the non-reviewing group recall performance was examined in this section. Table 4-19 lists the correlation between the recall performance and note contents of the non-reviewing group.

Table 4-19: The Correlation between Note Contents and Recall Task

Criteria	<i>r</i>	<i>p-value</i>
TW	0.34	0.13
IU	0.17	0.47
TA	0.12	0.42
EF	-0.17	0.46
CP	0.17	0.47
MI	0.12	0.60

Based upon the results, none of the note content criteria appeared significantly

correlated with the subjects' recall performance. Unlike the LC task, more of IU, TA, CP, or MI did not guarantee enhancement of the subjects' recall performance. Without the aids of notes, relying solely on their own memory, the subjects failed to establish a connection with their notes in the recall performance, indicating the importance of one of the note-taking functions – external storage.

Table 4-20 addresses the subjects' retention of the lecture on the recall task administered one week later.

Table 4-20: Retention of Information Units

Before Recall						Recall					
Major Information		Minor Information		Total		Major Information		Minor Information		Total	
Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
17.95	4.31	29.52	7.60	47.48	11.91	6.52	2.58	2.10	1.25	8.61	3.83

Before the recall, the subjects averagely recorded 47.48 information units while one week later, their memory of the lecture shrunk to 8.61 information units with 81.87% of memory loss, as shown in Table 4-21.

Table 4-21: The Percentage of Loss and Retention in the Recall Task

Delayed Recall					
Major Information		Minor Information		Total	
Loss	Retention	Loss	Retention	Loss	Retention
63.68%	36.32%	92.89%	7.11%	81.87%	18.13%

A detailed analysis into two types of information yielded that the retention of major information outnumbered that of minor information. The subjects were able to retain 36.32% of the recorded major information while they retained only 7.11% of the minor information with 92.89% of the minor information trimmed.

The 81.87% of memory loss and 18.13% of memory retention in the present study corresponded to Spizer (1939), who conducted an experiment to explore the relationship between retention and forgetting. In his study, it was shown that half of what the subjects heard would not be able to retain in their memory for more than one day unless they adopted the strategy of note-taking and reviewed their notes. As time went by, 2 weeks later, only 20% of the information would remain in their memory.

The results of the delayed recall in the non-reviewing group exhibited agreement with what Chaudron, Loschky, & Cook (1988) have found in their immediate recall. They stated that in terms of either the quality or quantity of notes, few significant correlations with the immediate reconstruction of the lecture. They explained that the immediate recall allowed their subjects to rely heavily on their short-term memory rather than their notes. Therefore, the subjects' notes could not exert their effect under such a condition. The delayed recall in the present study deprived the non-reviewing group of the chance refer back to their notes. Hence, the only tool that could serve as their reminder was the retention of their memory, which however failed to establish any correlation with their note contents.

4.5 Summary of Chapter 4

After examining the six criteria, in relation to the subjects' test performance, test-answerability served best to indicate their performance. In addition to test-answerability, major information, information units, and completeness all exerted different degrees of correlation with the subjects' test performance. With regard to the

recall task, the chance of reviewing notes contributed to a greater number of information units recalled. Only under the note-reviewing circumstance did the examined criteria except efficiency signaled some correlations with the subjects' recall performance. Deprivation of the chance to review the notes resulted in the subjects' fewer information units reported in the recall. In addition, global and local questions displayed different relationships with the subjects' note content. Global questions did not establish any correlation with any of the criteria while local questions had similar results like the subjects' overall test performance. Test-answerability held the strongest correlation, followed by major information, information units, and completeness in turn.