The Reading Matrix
Vol. 6, No. 1, April 2006

PASSAGE DEPENDENCY OF READING COMPREHENSION ITEMS IN THE GEPT AND THE TOEFL

Shiauping Tian
sptian@mail.ntust.edu.tw

Abstract

This study investigates the passage dependency of selected reading comprehension items from the GEPT (General English Proficiency Test) and the TOEFL (Test of English as a Foreign Language) and examines students' responses to items with extremely low passage dependency. Thirty-seven reading comprehension items selected from the two tests were administered to a group of ninety-three university students in both passage-out and passage-in conditions. Results of passage dependency analysis are presented according to test and item type. The selected items from the two tests do not differ significantly in passage dependency index, although the average passage dependency index of items from the GEPT is slightly higher than that of items from the TOEFL. Also, items about details appear to be more passage dependent than inference items. Three items with zero or negative passage dependency are identified, all vocabulary items from the TOEFL, indicating that the passage might have in some way misled the students. The paper then presents detailed examination of these three vocabulary items along with results of further tests to identify possible sources of confusion. The author concludes with implications for instruction and future studies.

Introduction

The knowledge that readers bring to written texts is crucial in comprehension process, as comprehension calls for interaction of previous knowledge with new information. The results of reading comprehension tests are considered to reflect a combination of two kinds of knowledge—readers' previous knowledge and information gained from reading the texts. Therefore, researchers have developed an interest in the extent to which readers are able to correctly answer reading comprehension questions without the passage on which the questions are based. Such information is considered key to the validity of reading comprehension items. Researchers have generally agreed that since reading comprehension tests are

designed to measure readers' ability to make sense of written texts, questions that can be answered correctly without the accompanying passage would call into question the validity of such reading tests. As stated by Tuinman (1973-1974, p. 221), "everything else being equal, the test with the most items with the highest degree of passage dependency offers the largest guarantee against invalidity due to responding to items without prior reading of the passage on which the item is based." Passage dependency indicates the extent to which reading comprehension items require the presence of the stimulus passage to be answered correctly.

Due to the various amount of pre-knowledge that readers possess and test-taking strategies that readers can apply, it is extremely difficult to construct reading comprehension items that can be answered correctly only with information from the stimulus passage. Nevertheless, researchers believe that if test-takers can score significantly higher than chance level on a reading test without access to the stimulus passages, then it is highly possible that the test is actually measuring something other than comprehension because factors other than the ability to understand the passages contribute significantly to variation of scores.

Related Literature

The issue of passage contribution or passage dependency of reading comprehension items was discussed extensively in the sixties and seventies. Attempts were repeatedly made to call attention to the need for test-takers to read the stimulus passage in order to successfully answer comprehension items (Johns, 1978; Pyrczak, 1975; Pyrczak & Axelrod, 1976; Tuinman, 1973-1974). Several studies revealed that students often managed to answer reading comprehension questions correctly even when the associated passages were not available (Preston, 1964; Pyrczak, 1972, 1974, 1975), and such findings were viewed as a potential threat to the validity of reading comprehension tests. The underlying assumption was that what's measured with the passages removed was something other than comprehension, and low correlation of students' performance in the passage-removed and the conventional conditions have been used as evidence that indeed two types of behaviors were measured in the two conditions (Hanna & Oaster, 1978; Preston, 1964; Tuinman, 1971). To discuss the issue of passage dependency in a more systematic way, Tuinmann (1973-74) proposed ways to measure the degree to which reading questions could be answered without reading the stimulus passage by calculating the passage dependency index for each question. In his study, data on five major standardized reading tests were obtained from over 9,000 4-6 graders in total. Results of his study showed that the passage dependency for all five tests was relatively low. On one test the subjects were even able to get half of the items correct without access to the passages. Another important

finding was that the degree to which test items were passage dependent was related to the age of the test-takers. In other words, the older the test-takers, the more sophisticated they were in terms of extracting clues to arrive at correct responses without the help of the passage.

Perkins and Jones (1985) investigated the passage dependency of two commercial reading tests by administering the tests in two conditions, with and without the stimulus passages, to 44 international students (freshmen and sophomores) in a university with an interval of three weeks. The calculated passage dependency index was negative for both tests, and the authors therefore stated that for the particular pool of subjects, the majority of items on the two tests were assessing background knowledge that was not gained through reading the passages. The subjects' performance in the two conditions was also analyzed according to the types of items contained in the two tests, and the results showed no significant difference for any of the item type under the two conditions. In addition, results of latent trait measurement indicated that for these two tests, items that required inferential and paraphrasing skills appeared to be the most difficult for the subjects.

Katz et al. (1990) designed two experiments in which university students were required to answer reading comprehension items on the SAT with and without the passages. They found that students' scores on the SAT verbal section correlated highly with their performance in both passage-out and passage-in conditions. In other words, variability of the students' performance on the reading comprehension items in both passage-in and passage-out conditions could be accounted for by their SAT-V scores. The authors therefore concluded that performance on the reading comprehension items of SAT seemed to depend substantially on factors unrelated to comprehending the accompanying passages.

To further pinpoint factors involved in the item answering process without the passages, Katz et al. (1991) intentionally removed "cognates" (information from other items of the same passage) in the passage-out condition and found that, despite the change, students were still able to score well above chance level without the passages. Therefore, it was inferred that when the passages were unavailable, students relied more on outside knowledge, whether knowledge about the subject matter or cultural knowledge, and test-taking skills rather than information from other items belonging to the same reading passage to get the correct answers. In a later study, Katz and Lautenschlager (2001) used regression analysis to assess the roles of passage and non-passage related factors in variance of item difficulty on SAT reading comprehension tasks. Their results showed that non-passage factors were more important in accounting for variance in item difficulty and total variance.

In another study by Powers and Wilson (1993), passage dependency of SAT

reading comprehension questions and, particularly, the strategies used by students to achieve better-than-chance performance on the test without access to the passages were examined. 350 secondary school students were asked to answer 18 reading comprehension questions taken from three forms of the SAT and afterward to complete a questionnaire about the strategies they had used. Results of their study showed that the students were indeed able to score above chance level without the passages. In terms of the link between strategy use and performance, strategies involving verbal reasoning were found to be more productive than those based on characteristics of the questions or answer choices. The authors concluded that while the selected items did not seem to depend exclusively on information in the reading passages, the importance of factors not related to the passage was relatively limited. In addition, the other factors that impacted most heavily on test performance without the passages involved verbal reasoning skills. The authors commented that verbal reasoning abilities were actually important in students' academic success, and that the reasoning strategies students used in the passage-out condition were not as totally irrelevant to reading comprehension as viewed by others, since flexibility in strategy use was also one of the important traits demonstrated by good readers.

Method

Participants

Ninety-three students in a university in Taiwan participated in the main study. These students were all non-English majors in their second to fourth year in the university. Eighty students (thirty-two English majors and forty-eight non-English majors) other than the original group of participants took additional tests in the later stages of the study.

Procedure

The test included four passages from GEPT (General English Proficiency Test) reading comprehension (15items) and two passages from TOEFL reading comprehension (22 items). The GEPT was developed by the Language Training and Testing Center in Taiwan. Scores on the GEPT have been used as a reference of language proficiency by more than 50 organizations and post-secondary schools in Taiwan since its first administration in 2000. The GEPT is a criterion-referenced test with five levels (Elementary, Intermediate, High-Intermediate, Advanced, and Superior) that emphasizes all four skills. The test content is not based on any particular curriculum or teaching material and is expected to reflect features of the local culture. The level chosen for this study is the High-Intermediate (equivalent to non-English major college graduates). These students were first given the test in the

passage-out condition and told to indicate their reasoning strategies on the space provided right next to each item. To minimize memory effect, the students were given the test again in the passage-in condition four weeks after the first test.

Data Analysis

Contribution of the passages to the selected items was calculated using an index of passage dependency proposed by Tuinman (1973-74), and the results are presented by test and item type. Tests of significance were performed using SPSS software (version 11). For some items with particularly low passage dependency, additional tests were then designed and administered to groups of students different from the one taking the original test in an attempt to provide explanations or pinpoint possible sources of confusion.

Results and Discussion

Passage Dependency of GEPT and TOEFL Reading Comprehension Items

The items included in the tests were categorized with reference to Rosenshine's (1980) taxonomy of skill hierarchies. Two subcategories (identifying referents of pronouns and recognizing author's attitude/purpose) were added to accommodate the items appearing in the selected passages. Table 1 shows item distribution of the two tests.

Tables 1 Distribution of Item Types in Each Test

Item Type	Number of items	
	GEPT	TOEFL
Locating details		
Recognition		1
Paraphrasing/matching	5	5
Subtotal	5	6
Simple inferential skills		
Understanding words in context		8
Recognizing the sequence of events		1
Recognizing cause and effect	2	2
Comparing and contrasting	3	
Identifying referents of pronouns		2
Subtotal	5	13
Complex inferential skills		
Recognizing the main idea/title/topic	3	2
Drawing conclusions	1	

Predicting outcomes		
Recognizing the author's purpose/attitude	1	1
Subtotal	5	3
Total	15	22

Participants' performance on the tests is summarized in Table 2. In the views of Pyrczak and Axelrod (1976), if test takers score significantly higher than chance score (number of items divided by number of choices in each item), then something other than comprehension is being measured. Table 2 shows that the participants scored significantly above chance level (3.75 for the GEPT and 5.5 for the TOEFL) for both tests ($p<0.005$) in the passage-out condition.

Table 2 Mean Scores of Each Test and Item Type

	GEPT		TOEFL	
Item type	Passage-out	Passage-in	Passage-out	Passage-in
Locating details	0.76 (15.2%)	2.67 (53.4%)	1.64 (27.3%)	2.60 (43.3%)
Simple inferential skills	1.63 (32.7%)	2.57 (51.4%)	4.35 (33.5%)	5.45 (41.9%)
Complex inferential skills	1.94 (38.8%)	2.74 (54.8%)	1.22 (40.7%)	1.55 (40.7%)
Total	4.37 (29.13%)	8.03 (53.53%)	7.48 (34%)	9.95 (45.23%)

Two types of indexes were found in related literature to quantify contribution of the passage to the process of answering the comprehension questions. Pyrczak (1972) proposed an index of passage dependency:

$I=P_e-P_m$, in which

I= extent to which a comprehension item can be answered correctly in the absence of the reading passage,

P_e= proportion of examinees who are expected to answer correctly on the basis of chance alone, and

P_m= proportion of examinees who actually answer correctly.

Tuinman (1973-74) proposed another passage dependency index:

PDI=1-(d_{np}/d_p), in which

PDI= passage dependency index for a comprehension item,

d_{np} = proportion of correct responses to an item under the passage-out condition, and

d_p = proportion of correct responses to an item under the passage-in condition.

For the current study, Tuinman's PDI was adopted because it takes into account the relationship between the proportions of correct responses under the two conditions. A larger PDI indicates a better item in terms of passage dependency.

The passage dependency index for each item type of the two tests is presented in Table 3. For the selected passages, items in the GEPT seem to be more passage dependent than those in the TOEFL, although the difference is not statistically significant (p=0.062). Overall, the passage seems to contribute more to responses to items about details than inference items (p<0.05).

Table 3 Mean PDI for Each Test and Item Type

Item type	GEPT	TOEFL
Locating details	0.55	0.42
Simple inferential skills	0.40	0.27
Complex inferential skills	0.30	0.27
Average of total	0.47	0.31

The PDI of individual items reveals seven items (two from the GEPT and five from the TOEFL) as particularly low in passage dependency (PDI ≤0.20). Table 4 shows the distribution of these low PDI items by item types. Responses to these items are presented in Tables 5 and 6. From Table 4, we can see that four out of the seven low PDI items are vocabulary items.

Table 4 Distribution of Low PDI Items

	Item Number	
Item type	GEPT	TOEFL
Locating details		
Recognition		29
Paraphrasing/matching		
Simple inferential skills		
Understanding words in context		12,14,22,28
Recognizing the sequence of events		
Recognizing cause and effect	36	
Comparing and contrasting		

	Identifying referents of pronouns	
Complex inferential skills		
	Recognizing the main idea/title/topic	40
	Drawing conclusions	
	Predicting outcomes	
	Recognizing the author's purpose/attitude	

Table 5 Student Responses to Low PDI Items in the GEPT

	#36		#40	
Choice	P/O[1]	P/I[2]	P/O	P/I
A	9	11	12	5
B	12	10	*62*	*70*
C	8	6	14	11
D	*64*	*66*	5	7
PDI	0.03		0.11	

1. P/O= Passage-out condition
2. P/I= Passage-in condition

Table 6 Student Responses to Low PDI Items in the TOEFL

	#12		#14		#22		#28		#29	
Choice	P/O	P/I	P/O	P/I	P/O	P/I	P/O	P/I	P/O	P/I
A	6	14	9	33	10	8	17	29	13	7
B	24	18	18	12	*63*	*66*	3	4	12	5
C	*35*	*32*	*46*	*38*	14	12	37	24	21	25
D	28	29	20	10	6	7	*36*	*36*	*47*	*56*
PDI	-0.09		-0.21		0.05		0		0.16	

One of these low PDI items (#28 in the TOEFL) has a PDI of zero. Two of these items have negative PDI values (#12 and #14 in the TOEFL), meaning the item difficulty indexes decrease from the passage-out condition to the passage-in condition; in other words, more students chose an incorrect response after reading the passage on which the item was based. This represents an undesirable condition in which the presence of the passage reduces the chances of getting a correct response or prevents the respondents from showing what they know, suggesting that the passage may be confusing or misleading to the respondents.

In addition, from the figures above, some choices stand out. Three distractors appear to be much more appealing to students in the passage-in condition—distractor

(A) in item #12, (A) in item #14, and (A) in item #28. The following will discuss these zero or negative PDI items in detail.

Zero or Negative PDI Items

It's worth noting that all three items are vocabulary-defining items. On the surface, it seems quite reasonable for vocabulary items to have low PDIs. It is highly possible for some respondents to be familiar with the definition of the vocabulary tested, and the prototypical meaning of the vocabulary might also be conveyed in the context provided by the passage. The word tested in item #28, *nevertheless*, could well fit the description. Item #28 has a PDI of zero, indicating that the presence of the passage made no contribution to the respondents' reasoning process, which appears to be expected considering the low difficulty of the word (two stars on the three-star scale of word frequency used in *Macmillan English Dictionary for Advanced Learners of American English*).

However, if we examine the figures in Table 6, we can easily see that for some reason, the passage led a lot of students to interpret the word in a very unusual sense—(A) *therefore*. Before any further discussion is made, we should first take a closer look at these three vocabulary items.

Despite the lack of unanimous opinions among researchers regarding the inclusion of vocabulary items in reading comprehension tests, figuring out meanings of unknown words from context is generally considered an essential skill in successful comprehension, and some guidelines were provided about construction of vocabulary items in reading comprehension tests. Hill & Larsen (2000) proposed three guidelines for constructing a vocabulary-defining item:

(1) The vocabulary-defining function should be made explicit.
(2) The vocabulary item to be defined should be one that [test-takers] are not likely to know.
(3) The context of use should allow [test-takers] to make sense of the vocabulary item. (p.374)

Evaluation of the three items according to these guidelines is summarized below.

Table 7 Evaluation of the Three Zero or Negative PDI Items

Guideline	#12	#14	#28
Explicitly-worded vocabulary-defining function	✓	✓	✓
Likely unknown vocabulary item	✓	✓	
Context allows sense making of vocabulary item	✓	✓	✓

Table 7 shows that all three vocabulary items meet the first guideline by making the vocabulary defining intent explicit to test-takers, and two of them meet the second guideline by testing vocabulary items likely to be unfamiliar to test-takers. Results for the third guideline were marked in gray for two reasons. The negative PDIs for items #12 and #14 suggest that the presence of context not only failed to help the respondents make sense of the vocabulary item, it even hurt their chances of making the right choice by leading more people to a particular distractor. The zero PDI for item #28 also indicates that the context made no contribution to the sense making process. From the evidence of PDI alone, it seems that none of the three items meets the third guideline. However, mature readers might have different judgments in this regard. To go beyond what's revealed by PDI, the following will discuss in detail the role of context in the meanings of the three words in question and also why the passage rendered certain distractors a lot more plausible.

Item #12
The word "magnified" in line 6 is closest in meaning to
(A) caused
(B) modified
(C) intensified
(D) combined

This item referred to the section of passage below:

> In winter especially, it is important for birds to keep warm at night and conserve precious food reserves. One way to do this is to find a sheltered roost. Solitary roosters shelter in dense vegetation or enter a cavity—horned larks dig holes in the ground and ptarmigan burrow into snow banks—but the effect of sheltering is ***magnified*** by several birds huddling together in the roosts, as wrens, swift, brown creepers, bluebirds, and anis do. Body contact reduces the surface area exposed to the cold air, so the birds keep each other warm. Two kinglets huddling together were found to reduce their heat losses by a quarter and three together saved a third of their heat.

For mature readers, the context does provide some clues about the meaning of the vocabulary item *magnified*. The last sentence in the excerpt clearly states that two birds huddling together reduced heat losses by a quarter and three birds together saved a third of their heat. In other words, the more birds huddle together, the less heat loss and therefore the greater the effect of sheltering. However, the clues were not found in

the nearest context of the word tested (the sentence in which the vocabulary appeared in and that before or after it) but in the extended context. This, together with the possibility that the appearance of many unfamiliar names of birds near the end of the sentence might have frustrated the test-takers enough to prevent them from reading on to seek clues, could probably explain why so many test-takers failed to infer the meaning based on the clues provided. Distribution of responses in Table 8 shows that more students were attracted to distractor (A) in the passage-in condition. The reason the passage led more students to this distractor is probably that the appearance of *by* activated the students' knowledge of the familiar phrase *be caused by* and prompted them to choose a word substitutable for the immediate context.

As for the popular distractor, *(D) combined*, students who chose this distractor in the passage-in condition probably comprehended the message conveyed in the text—that if several birds huddle together, the effect of sheltering would be greater. Their problem seemed to lie in word choice. To these students, if several birds huddle together, the effect of sheltering would be "added up" and hence the choice *combined*. The researcher suspected that if (C) was changed to *strengthened* or *enhanced*, a more commonly recognized word, the distribution of responses would probably be rather different. The rationale was that if students could understand both (C) and (D), then those who comprehended the passage could make a comparison and find (C) a better choice. To test the idea, the researcher gave the modified item to another group of 48 non-English majors in the same university. These students were first told to answer the question without the passage and then, after their answers were collected, to answer the question again with the accompanying passage present. The students were also asked to write down the definition of *magnified* and explain their choice in the passage-out condition, and to summarize the meaning of the passage and explain their choice in the passage-in condition. For this test, only the section of passage shown above, instead of the entire original passage, was given in the passage-in condition. Results of the test are reported in Tables 8 and 9.

Table 8 Responses to Item #12

Choice	Original (N=93)		Modified (N=48)	
	P/O	P/I	P/O	P/I
A	6 (6.5%)	14(15.1%)	4 (8.3%)	3 (6.3%)
B	24(25.8%)	18(19.4%)	7 (14.8%)	12 (25%)
C	*35(37.6%)*	*32(34.4%)*	*30 (62.5%)*	*17 (35.4%)*
D	28(30.1%)	29(31.2%)	7 (14.6%)	16 (33.3%)

Table 9 Analysis of Responses to Item #12 (Modified Version)

P/O	P/I				
	A	B	*C*	D	Total
A	0	0	1	3	4
B	0	1	0	6	7
C	*3*	*8*	*13*	*6*	*30*
D	0	3	3	1	7
Total	3	12	17	16	48

Results of the test failed to support the researcher's hypothesis. Substituting *intensified* with an easier word that the students were more likely to recognize seemed to make the choice even less appealing in the passage-in condition. On the other hand, the modified version seemed to make, unexpectedly, the popular distractor (D) even more attractive to the students. A few possible reasons were gathered from responses by students who cared to explain their choices. The first explanation is that the number of students choosing (C) in the passage-out condition is quite inflated. According to the students' responses, four of them who could not comprehend the sentences decided to pick (C) for no apparent reason. The students seemed to be under the impression that (C) would be the best guess when they came across an unanswerable item. In addition, seven other students simply went for the longest choice in the passage-out condition, probably due to past training on test-taking strategies, because they had no idea what *magnified* meant. Although the correct response for this item is indeed the longest among the four choices (6 letters in choice A, 8 in B and D, 11 in C, and 12 in C of the modified version), the difference is hardly great enough to constitute an additional clue. Therefore, at least 11 out of the 30 students chose the right answer for the wrong reason in the passage-out condition, and it's quite possible that the students in the original test applied the same kind of reasoning.

As for those who did get the message conveyed in the few sentences, quite a few of them, at least 5 based on the responses, were attracted to *combined* due to partial knowledge of word definition and how the word should be used in context.

These students did not recognize that substituting *magnified* with *combined* in the sentence did not fit the use of *combine* in the sense of joining two things to become one, or the expression of "one thing being combined with another". They were probably influenced by the Chinese translation of the word and interpreted *combined* as "put together" and therefore "added up". Since some students who could comprehend the message were led to distractor (D), the question of whether the

distractor worked against the better readers naturally came to mind. From the responses of the top and bottom 27% of the group, distractor (D), unlike (A) and (B), appeared to be an adequate distractor, distracting a lot more students from the bottom than the top group (See Table 10).

Table 10 Responses to Item #12 from High and Low Group

	A	B	*C*	D
H (N=25)	4	5	*11*	5
L (N=25)	5	5	*3*	12

Item #14
The word “forage” in line 12 is closest in meaning to
(A) fly
(B) assemble
(C) feed
(D) rest

Item #14 referred to the section of passage below:

> During the day, parties of birds will have spread out to ***forage*** over a very large area. When they return in the evening some will have fed well, but others may have found little to eat. Some investigators have observed that when the birds set out again next morning, those birds that did not feed well on the previous day appear to follow those that did.

For this item, clues exist in phrases like *have fed well* and *have found little to eat*. From these words, test-takers could infer that the purpose of the birds’ trip during the day had something to do with food. More clues could also be found in the more extended context, since the following few sentences mainly talk about how different kinds of birds have different hunting habits. For this item, distractor (A) appears to be particularly attractive in the passage-in condition. The popularity of this distractor can be reasonably attributed to the association of *birds* with *fly*, which would not have been so strong without the presence of context. Another reason is substitutability of the choices. Among the four words in the choices, only *fly* and *feed* can be substituted in the original context to form a logical sentence. There is little surprise then that with the appearance of *birds* in the context, a lot more students were prompted to choose (A) in the passage-in condition.

As in item #12, many participants failed to choose the target response because they only focused on the single sentence in which the word appeared and did not

bother to read on to seek further clues.

Item #28
The word “Nevertheless” in line 24 is closest in meaning to
(A) therefore
(B) because
(C) occasionally
(D) however

This item referred to the section of passage below:

> Almost everyone now had a more diversified diet. Some people continued to eat mainly foods that were heavy in starches or carbohydrates, and not everyone could afford meat. ***Nevertheless***, many families could take advantage of previously unavailable fruits, vegetables, and dairy products to achieve more varied fare.

This section appears in the concluding paragraph of the stimulus passage. The previous few paragraphs mention the lack of variety in people’s diet in the past due to limited food preserving technologies and introduce the invention of various devices to preserve food in history. The author concludes that now that more advanced technologies are available to preserve food, it’s easier for people to achieve diversity in diet. A contrast is presented in the paragraph about people’s diet nowadays—for ***some people***, their diet is still limited in variety, but for ***many families***, they can take advantage of the technologies to achieve diversity in diet.

Before looking at the response distribution, the researcher expected the word *nevertheless* to be known to many of the participants and expected a high rate of correct response in the passage-out condition. The figures in Table 6, therefore, came as a surprise in that only about one third of the participants recognized the word without its context, and that no more participants figured out its meaning even with the help of context. The fact that choice (C) *occasionally* became the most popular distractor in the passage-out condition can be easily explained. The students merely extracted *never*, a word they recognized, from the longer word *nevertheless* and then chose a word that also denoted frequency. Questions remain about why the context failed to clarify the relationship between these sentences, and also why the presence of the passage made distractor (A) *therefore* much more appealing to the students. Does this mean that the passage somehow misled the students in a certain way?

In an attempt to answer the questions, the researcher asked a group of 32 English majors in the university to answer this item in both conditions to see if the passage

could really lead students, even those who knew the meaning of *nevertheless*, to misinterpret it. The 32 students were asked to first answer this item in the passage-out condition and then the passage-in condition.

Table 11 Responses to Item #28 by English Majors

Choice	P/O	P/I
A	0 (0%)	5 (16%)
B	0 (0%)	0 (0%)
C	1 (3%)	1 (3%)
D	*31 (97%)*	*26 (81%)*

Responses in the passage-out condition showed that almost all of these English majors understood the word *nevertheless* without the help of context. Results in the passage-in condition, however, indicated that even to readers who understood the word, the passage did lead some of them to misinterpret it.

With (A) being the most popular distractor, it can be seen that students who picked the wrong choice tended to view the last two sentences as presenting a cause-effect relationship instead of a contrast. More specifically, they interpreted the relationship as problem-solution. A closer look at the original text may render the cause obvious. The first half of the first sentence states that some people ate mostly foods rich in starches and carbohydrates, but the reason for which is not mentioned. The later half of the sentence explains that some people didn't have a diversified diet because they could not afford meat. Then the next sentence states that many people could have a diversified diet by having fruits, vegetables, and dairy products, thanks to modern technology. The students simply picked up the later half of the first sentence, that some people couldn't afford to eat meat, and viewed fruits, vegetables and dairy products mentioned in the second sentence as a solution offered by modern technology to those who couldn't afford meat. The juxtaposition of *meat, fruit,* and *vegetables* was particularly distracting because of the common reasoning that if you couldn't eat meat, then you just ate more fruits and vegetables now that modern devices have made them easily available. Hence, for the few students, the contrast signaled by *some people* and *many families* was outweighed by the common association prompted by *meat*, *fruits*, and *vegetables*.

If students who knew the meaning of *nevertheless* could be misled by the context, then it was little surprise that students who did not know the word were tempted to misinterpret the sentences as presenting a cause-effect relationship. Even so, the signals of contrast provided by the text seemed to be clear enough, considering the overall message that almost everyone had a diversified diet nowadays. Also, students

are not supposed to interpret the passage based solely on associations prompted by bits and pieces of information in the texts and ignore the logic inherent in the text. For example, if the cause-effect relationship truly exists between the two sentences, then the last sentence would most likely be phrased as “Therefore, many families *took* advantage of…” to denote a fact instead of “Therefore, many families *could take* advantage of…” to express a possibility.

Conclusion

Passage dependency is the extent to which answers to reading comprehension questions depend on information in the reading passages. Since a reading question that can be answered without the stimulus passage can hardly be called a measure of reading comprehension, passage dependency has been used as evidence against potential invalidity of reading comprehension tests. The present study examines items from reading comprehension sections of two frequently used tests in Taiwan—the GEPT and the TOEFL. Items with zero or negative passage dependency were analyzed for possible sources of confusion or misinterpretation. Results of the study show that for the items selected, no significant difference in passage dependency was found between items from the two tests, although items from the GEPT had on average a higher passage dependency index than those from the TOEFL. Vocabulary items from the TOEFL were found to be particularly low in passage dependency, which many may be tempted to explain as a result of the test-takers’ pre-knowledge of the tested vocabulary items. However, based on the guidelines for writing vocabulary-defining items by Hill & Larsen (2000), passage dependency for this type of items should be expected to be high, since the vocabulary in question should be one that the test-takers are unlikely to know and whose meaning can be derived from the context.

Examination of three zero or negative passage dependency items points to three possible causes for such results: false belief about vocabulary-defining items on these tests, inadequate knowledge of word usage, and inappropriate use of real world knowledge to interpret texts. Based on the analysis, the following recommendations can be made regarding future reading instruction. First, it is insufficient to introduce a new word with its Chinese translation only. An English word and its correspondent Chinese term might not be exactly equivalent. If students are given the Chinese translation only, then they might be prone to errors due to subtle differences on both semantic and syntactic levels. In addition to word definition, more efforts should be made on word usage in vocabulary instruction. For the students, being aware of how to use a word in context is as essential as knowing the meaning of the word. In this respect, instructors should not feel confined to the limited example sentences provided

by textbooks or some well-established dictionaries. The Internet and concordancing software can be a good source of authentic examples reflecting how the word is used in real contexts. Other important aspects instructors need to focus on include vocabulary skills and appropriate use of real world knowledge to assist comprehension. Students need to receive direct instruction and sufficient practice on dealing with unknown vocabulary in contexts. Sufficient practice in this regard could help students break free from the habit of engaging in a word-by-word decoding process when reading and keep them from panicking in the face of unfamiliar words. In addition, studies on skilled and unskilled readers have shown that more proficient readers tend to use real world knowledge to assist comprehension or confirm hypothesis formed in the reading process, while less proficient readers tend to be distracted by associations unrelated to the passage. Therefore, students probably need to be instructed on using real world knowledge to support the larger framework of overall meaning so that they would not allow associations derived from real world knowledge to lead them away from the texts and end up losing sight of the larger meaningful units. A minor point to mention is that it's probably worthwhile for instructors to find out about students' false beliefs about taking reading tests of this type. For example, some students might be under the impression that for vocabulary-defining items, all they need to do is to read the sentence in which the word appears while the truth is that clues may exist not in the immediate context but in more expanded contexts.

Finally, future studies on similar topics can take limitations of the present study into consideration. To begin with, this study used a rather small number of items from each of the two tests. A much larger number of items from each test need to be included for any general claim to be possible about any test. Next, although the study attempted to elicit the participants' reasoning processes during the passage-out condition by asking them to write down the rationale for each response, many of the participants failed to provide any explanation for most of the items. Future studies can be conducted with an improved design to collect more data about the participants' thinking processes in both passage-out and passage-in conditions. Only then would analysis of the low PDI items be more thorough and capable of providing more practical implications for item construction and reading instruction. In addition, the usual practice of using the same items and passages on the same group of participants in the two conditions is believed to introduce factors that might distort the results. A more ideal design would be to use alternate forms for each test, as in the study by Hanna and Oaster (1978), if such forms are available.

References

Hanna, G., & Oaster, T. R. (1978). How important is passage dependence in reading comprehension? *The Journal of Educational Research, 71*(6), 345-348.

Hill, C. A., & Larsen, E. (2000). *Children and Reading Tests*. Connecticut: Ablex.

Johns, J. L. (1978). Do comprehension items really test reading? Sometimes! *Journal of Reading, 21*(7), 615-619.

Katz, S., & Lautenschlager, G. J. (2001). The contribution of passage and non-passage factors to item performance on the SAT reading task. *Educational Assessment, 7* (2), 165-176.

Katz, S., Lautenschlager, G. J, Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1*, 122-127.

Katz, S., Blackburn, A. B., and Lautenschlager, G. J. (1991). Answering reading comprehension items without passages on the SAT when items are quasi-randomized. *Educational and Psychological Measurement, 51*, 747-754.

Perkins K. & Jones, B. (1985). Measuring passage contribution in ESL reading comprehension. *TESOL Quarterly, 19*(1), 137-153.

Powers, D. E., & Wilson, S. T. (1993). Passage dependence of the new SAT reading comprehension questions. *College Board Report*, 93-3.

Preston, R. C. (1964). Ability of students to identify correct responses before reading. *The Journal of Educational Research, 58,* 181-183.

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice items designed to measure comprehension of reading passages. *Reading Research Quarterly, 8*, 62-71.

Pyrczak, F. (1974). Passage-dependence of items designed to measure the ability to identify the main ideas of paragraphs: Implications for validity. *Educational and Psychological Measurement, 34*, 343-348.

Pyrczak, F. (1975). Passage-dependence of reading comprehension questions: Examples. *Journal of Reading, 18,* 308-311.

Pyrczak, F., & Axelrod, J. (1976). Determining the passage dependence of reading comprehension exercises: A call for replications. *Journal of Reading*, *19*(4), 279-283.
Rosenshine, B. V. (1980). Skills hierarchies in reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.). *Theoretical issues in reading comprehension* (pp.535-554). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Tuinman, J. J. (1973-74). Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, *9*(2), 206-223.

Macmillan English dictionary for advanced learners of American English. (2002). Hampshire, England: Macmillan.

Dr. Shiauping Tian is an assistant professor in the Department of Applied Foreign Languages at National Taiwan University of Science and Technology. Her research interests include EFL reading strategies, assessment of reading, and computer-assisted language teaching. Address: Dept. of Applied Foreign Languages, National Taiwan University of Science and Technology, No.43, Sec. 4, Keelung Rd., Taipei, Taiwan (106) Email: sptian@mail.ntust.edu.tw