

Factors Affecting the Difficulty of Short Dialogue Listening Items of Test Of Proficiency-Huayu (TOP-Huayu)

Introduction

Test of Proficiency-Huayu (TOP) is developed by the Steering Committee for the Test of Proficiency-Huayu (SC-TOP) and has administered worldwide since 2003. Up to now, SC-TOP has received many suggestions on the test content. One of the suggestions is about the item difficulty, which, if not effectively controlled, might reduce the power in discriminating less proficient test takers from more proficient test takers. That is, a test with too many easy or difficult items may cause ceiling effect and floor effect, respectively, which will weaken the discrimination power. In fact, an ideal language test should contain a balanced number of test items with various levels of difficulties so that the discrimination power can be maximized. From this point of view, it is of great importance to control the item difficulty in a test. Moreover, in the item-writing guideline developed by SC-TOP, the issue of adjusting the item difficulty is less touched upon when being compared to other dimensions, such as topic selection, task type design, determination of abilities to be measured and so forth. Therefore, the purpose of the present study is to find out the potential factors that may affect the item difficulty, and further to provide item writers some useful principles in controlling the difficulty of an item.

This study, following the format adopted in a TOEFL research report (Kostin, 2004), attempts to identify linguistic factors in the test itself, as well as test takers' task-processing factors, which may affect the test the item difficulty in the short dialogue listening section of TOP-Huayu for Advanced Level. A correlation analysis and a regression analysis will be conducted to interpret the data.

Method

Materials

A total of thirty-five short-dialogue items from two versions of TOP pre-tests for Advanced Level were employed as the materials in this study. The two versions of tests were treated with the *common-item nonequivalent groups design*; thus, their compatibility of difficulty can be ensured. Each item contains a short dialogue involving one turn-taking of speaking, accompanied by a question about the content of the dialogue. Four options—one key and three distracters—were provided for the test takers to choose from.

Independent Variables

A total of 17 independent variables were selected from a TOEFL monograph

(Kostin, 2004) and modified (if necessary) to better reflect the characteristics of Mandarin Chinese. The 17 variables can be roughly divided into three categories, which are word-level variables, sentence-level variables and task-processing variables. Each variable, along with its coding instructions, is explained and exemplified in the following.

Word-level Variables

V01: The necessity of comprehending four-character idioms in correctly responding to the test items.

In English, an idiom is defined as “an expression consisting of two or more words having a meaning that cannot be deduced from the meanings of its constituent parts” (*The American Heritage Dictionary*, 2000, p. xxxvi). In Chinese, an idiom refers to a fixed colloquial expression (usually with a verb-object construction) which is characterized by humor, wittiness, and funniness. It is often meant to reflect certain social phenomena and behaviors (Shao, 2006). For example, “開夜車” is used to describe the behavior of working very late into the night. Like English idioms, the meaning of a Chinese idiom cannot be derived from its literal meaning; rather, it is historical, allegorical or metaphorical in nature. Since the meaning of a Chinese idiom is not the combination of the meanings of its constituent parts, it is predicted that the use of an idiom may make a test item more difficult.

The example of V01 is given below:

男：許教授的修養很好，非常樂意傾聽學生的意見。

女：我也聽說他不擺架子，常跟學生打成一片呢！

男：許教授為人如何？

Coding instructions for V02: Code “1” for the test item if the comprehension of the colloquial idiom is necessary for correctly responding to the item; otherwise, Code “0”.

Sentence-level Variables

V02: Two or more negatives in utterances of first speaker.

The influence of negatives in listening comprehension has been noticed in a few studies (e.g., Kostin, 2004; Nissan, DeVincenzi, and Tang, 1996). These studies found a positive correlation between the presence of negatives and the item difficulty; i.e., an item with two or more negatives is often more difficult than that with fewer negatives. In the present study, we adopted a more stringent negative-coding stipulation than Nissan (1996) did in her study. In Nissan’s (1996) study, negatives, be they negative or positive (e.g., tag questions) in meaning, were all counted. In the

present study, only the negatives (e.g., 不, 沒, 別, 非, 無, etc.) whose function are denial were counted because this can ensure that the influence on the item difficulty, if any, is made by its ‘pure negativeness’. A negative with a negative form but without the denial function, such as the negative in a rhetorical question (e.g., “難道我[沒]跟你說嗎?”) or A-not-A question (e.g., “是[不]是?”), may render the results more difficult to interpret in that it is not easy to determine whether the linguistic context where the negative occurs or the negative *per se* causes the item difficulty. In the current study, neither the negative marker in a rhetorical question nor the negative marker in an A-not-A question was counted because it lacked the “denial” function even if it manifested itself as a negative marker in form. For example, “沒” as in “根本沒時間看書” was coded as a negative, whereas “不” as in “我[不是]早就告訴過妳了嗎?” and “不” as in “至於其他人會[不]會信以為真就不得而知了” were not counted. The example of Variable 02 is given below:

男：昨天報上登的那個消息[不]知是真是假，也[不]知有多少人會真的相信？

女：是啊！但是對我來說，沒有多大的影響，至於其他人會不會信以為真就不得而知了。

男：這位小姐的意思是什麼？

Coding instructions for V02: Code “1” for the test item with two or more negatives in the utterances of the first speaker; otherwise, Code “0”.

V03: Two or more negatives in utterances of second speaker.

The example of V03 is given below:

男：我不懂你為什麼得花那麼多時間看這家公司的資料！

女：現在求職的陷阱那麼多，找工作[不]能[不]謹慎啊！

男：這位小姐說找工作時要怎麼樣？

Coding instructions for V03: Code “1” for the test item with two or more negatives in the utterances of the second speaker; otherwise, Code “0”.

V04: Two or More Negatives in Total Dialogue.

The example of V04 is given below:

男：我不懂你為什麼得花那麼多時間看這家公司的資料！

女：現在求職的陷阱那麼多，找工作[不]能[不]謹慎啊！

男：這位小姐說找工作時要怎麼樣？

Coding instructions for V03: Code “1” for the test item with two or more negatives in the utterances of the second speaker; otherwise, Code “0”.

V05: Use of complex sentences in utterances of first speaker.

A complex sentence is composed of an independent clause and a dependent clause. There is a certain semantic relation, such as giving reasons (e.g., 因為...所以...), making inferences (e.g., 既然...那麼...), expressing supposition (e.g., 如果...就...), stating conditions (e.g., 除非...否則...), offering concessions (e.g., 雖然...但是...), clarifying time (e.g., 一...就...) and so forth. The two clauses are usually combined with each other by device of conjunctions; however, it is not uncommon to see a complex sentence without any conjunctions. For example, “時間不早了，我們走吧” is a complex sentence involving a casual relation even though it lacks an explicit connector. Given that a complex sentence has a more complicated syntactic structure and contains more ideas than a simple sentence (<http://users.chariot.net.au/~michaelc/complex.htm>), it is theoretically more difficult to comprehend a complex sentence, and its presence may increase the difficulty of a test item. In the present study, we coded for the test item with at least one complex sentence. The example of V05 is given below:

男：既然妳跟小王之間有誤解，為什麼不解釋清楚呢？

女：唉！有什麼好說的，我再怎麼說，都一樣。

男：這位小姐的意思是什麼？

Coding instructions for V05: Code “1” for the test item containing at least one complex sentence in the utterance of the first speaker; otherwise, Code “0”.

V06: Use of complex sentences in utterances of second speaker.

The example of V05 is given below:

男：回國以後找工作難不難？

女：一般的工作應該不難，可是如果想找這樣打從心裡喜歡的工作可不容易。

Coding instructions for V06: Code “1” for the test item containing at least one complex sentence in the utterance of the second speaker; otherwise, Code “0”.

V07: Use of complex sentences in total dialogue.

Coding instructions for V07: Code “1” for the test item containing at least one complex sentence in the total dialogue; otherwise, Code “0”.

V08: Number of within-clause referentials in total dialogue.

Research has shown that a text will be easier to read if the referential is replaced with its referent or noun phrase because readers do not have to figure out what the referential refers to (Abrahamsen and Shelton, 1989). In addition, Kostin (2004) found that the number of within-clause referentials was positively correlated with the item difficulty in the short dialogue comprehension. In other words, the more

referentials a test item contains, the more difficult the item is. In the present study, three types of referentials were coded: within-clause referentials, between-clause referentials and inter-speaker referentials. Cautions should be taken that the referentials coded in the present study were not limited to “personal pronouns”; rather, they could extend to “demonstrative pronouns” (e.g., 這, 那, 這邊, 那邊, etc.) as well. The example of V08 is given below:

男：李小姐，妳生病了，怎麼不在家休息呢？

女：我只不過是輕微的感冒而已，連一點藥都沒吃就好了。

Coding instructions for V08: Code the number of within-clause referentials in the total dialogue.

V09: Number of between-clause referentials in total dialogue.

The example of V09 is given below:

男：我昨天請小王幫我把那份報表交給經理，沒想到他竟然忘記了，害我被經理罵了一頓。

女：別生氣了！所謂人非聖賢孰能無過，出點錯難免，你就原諒他吧。

Coding instructions for V09: Code the number of between-clause referentials in the total dialogue.

V10: Number of inter-speaker referentials in total dialogue.

The example of V09 is given below:

男：林小姐真不給面子，每次請她吃飯，總是有藉口。

女：哎呀！王經理，你別為難她了，她的確另外有約了。

Coding instructions for V10: Code the number of inter-speaker referentials in the total dialogue.

V11: Use of rhetorical questions in utterances of second speaker.

A rhetorical question is “a question to which no answer is expected” (Crystal, 1992). Rhetorical questions often cause much difficulty for students learning Chinese as a foreign language (CFL) (Dong, 2000; Zhao, 2000). According to Zhao (2000), there are four main sources of such difficulty. To begin with, though appearing as a question form, a rhetorical question does not really require a definite answer. Second, the meaning of a rhetorical question is hard to grasp due to its high dependency on contexts. Third, most CFL learners are not familiar with the function of a rhetorical question, nor are they familiar with the time and way of using it. Fourth, the definition for a rhetorical question used in the current textbook is not complete, nor is it clear enough for CFL learners to master. Since learning a rhetorical question is not an easy task for CFL learners, the presence of a rhetorical question may contribute to the item

difficulty. The example of V11 is given below:

男：要是沒有你幫忙處理這件事，我就被老闆罵慘了！

女：幫這點忙算什麼？小李，你就別放在心上了！

Coding instructions for V11: Code “1” if the utterance of the second speaker contains any rhetorical question; otherwise, Code “0”.

Task-processing Variables

V12: Any of the three distracters has more words that overlapped with the words in the dialogue than does the key.

Several studies found that test takers sometimes resort to lexical overlap between words in the text and words in an item’s options in figuring out the answer to that item, especially when they do not understand the item quite well (LTTC, 1999). Kostin (2004) found that an item tends to be easier when lexical overlap between the text of the dialogue and the key is higher than that between the text of the dialogue and the distracters. On the other hand, an item with more lexical overlap in the distracters is likely to be more difficult. In the present study, we coded for the item that had a larger amount of lexical overlap between any distracter and the dialogue than between the key and the dialogue, an item predicted to be more difficult. There were two stipulations for coding lexical overlap. First, if a word was shared by the text, the key and the distracters, it would not be coded for lexical overlap. Second, if a word in the key or distracter and a word in the dialogue had the same meaning, but different word classes, the two words were not counted as lexical overlap. For example, “工作” as in “他的工作很危險” and “工作” as in “他每天工作三小時” should not be coded for lexical overlap because the former is a noun, while the latter is a verb even though they bear the same meaning “work.” The example of variable 12 is given below:

男：小華平時上課都在打瞌睡，作業也不按時交，這次能考上大學真是出乎意料之外啊！

女：是啊！看來我們對他得重新評估了。

男：這位小姐的意思是什麼？

(A)小華值得大家效法 (0 overlapping words) → distracter

(B)小華得再考一次大學 (2 overlapping words) → distracter

(C)不應該只看小華平時的表現 (0 overlapping words) → key

(D)小華根本沒有考上大學的實力 (2 overlapping words) → distracter

Coding instructions for V12: Code “1” if any of the three distracters have more overlapping words than does the key; otherwise, Code “0”.

V13: A word or phrase in the key has a near-synonym relation with the one in the last clause of the total dialogue.

It has been found that repetition of words has a facilitative effect on listening comprehension for less proficient L2 listeners, while other devices of paraphrasing, such as synonyms, play a very meager role in their listening comprehension (Chaudron, 1995). In addition, Chiang and Dunkel (1992) found that more proficient EFL learners have an advantage over less proficient EFL learners in listening to a lecture, of which information is paraphrased. Such an advantage enjoyed by more proficient EFL learners is attributable to their richer vocabulary knowledge. Taken together, a test item in which a word in the key is synonymous with the one in the dialogue can theoretically be used to distinguish more proficient learners from less proficient ones. In the present study, near-synonyms, rather than synonyms, were coded because true synonyms rarely exist (Saeed, 2002). Near-synonyms can be defined as a group of words that are very similar in their meanings, but may have different distributions. Consider “年紀” and “年齡”. They all mean “age,” but they differ in their collocations. For example, people often say “小小年紀”, but not “小小年齡”; “入學年齡” is acceptable, but not “入學年紀”. The example of V14 is given below:

男：經濟不景氣，公司裁了一批人了，聽說可能還會有下一步的大動作呢！

女：再這樣下去，我們可能都要另謀出路了。

男：這位小姐的意思是什麼？

- (A) 恐怕快要失業了
- (B) 已經找好新工作了
- (C) 工作絕對不會丟的
- (D) 失業了，要快點找工作

Coding instructions for V13: Code “1” if a word or phrase in the key has a near-synonym relation with the one in the last clause of the total dialogue; otherwise, Code “0”.

V14: Necessity of comprehending implicit statement in correctly corresponding to the test items.

Generally speaking, an implicit statement is more difficult to understand than an explicit statement in that test takers have to go beyond the literal meaning of the implicit statement, whereas the meaning of an explicit statement is rather straightforward. The report of GEPT pre-test revealed that less proficient listeners usually had problems figuring out the meaning of an implicit statement, while more proficient listeners could easily grasp the implicit meaning (LTTC, 1999). Similar

findings can be observed in Nissan et al.'s (1996) study. They found that an item testing implicit information was more difficult than the one testing explicit information. This is because test takers have to make a logical and reasonable inference before they can get the real meaning of the implicit information. The example of V15 is given below:

- 男：像你這麼有能力又願意吃苦的人真是難得！
女：算了吧！我們又不是昨天才認識的。
男：這位小姐的意思是什麼？

Coding instructions for V14: Code "1" if it is necessary to comprehend the implicit statement in correctly responding to the test item; otherwise, Code "0".

V15: The proposition of the key is opposite to the one of any of the three distracters.

Variables 15-17 were selected based on the SC-TOP researchers' findings that test items with two options having opposite propositions appeared to affect test takers' strategies for choosing the answer, which indirectly affected the item difficulty. To test if it is the case, we included Variables 15-17 in our study. The example of V15 is given below:

- 男：上星期期中考，你覺得題目怎麼樣？
女：每次的題目都不容易，這次更是難得離譜。
男：這次的考題怎麼樣？
(A) 難得很 →key
(B) 難易適中 →distracter
(C) 沒有上次難 →distracter
(D) 一點兒都不難 →distracter

Coding instructions for V15: Code "1" if the proposition of the key is opposite to the one of any of the three distracters; otherwise, Code "0".

V16: The propositions of any two of the three distracters are opposite.

The example of V16 is given below:

- 男：哇！我從來沒有見過那麼古老，建築那麼美麗又那麼熱鬧的地方！
女：小張，我不是早就告訴過你了嗎？現在你到底是見識到了！
(A) 建議小張多到處看看 →distracter
(B) 批評小張真的很沒見識 →distracter
(C) 稱讚小張對建築很有研究 →distracter
(D) 抱怨小張不早聽她的建議 →key

Coding instructions for V16: Code "1" if the propositions of any two of the three distracters are opposite; otherwise, Code "0".

V17: The propositions of any two of the four choices are opposite.

Please see V15 and V16 for the example of V17.

Coding instructions for V17: Code “1” if the propositions of any two of the four choices are opposite; otherwise, Code “0”.

Coding

Eight SC-TOP test researchers and developers participated in the coding work. There were three stages in the coding work: familiarization stage, coding stage, and agreement stage. At the familiarization stage, the author explained to the researchers the definitions of the 17 variables and the coding methods. At the coding stage, each researcher coded for the 35 test items independently. At the agreement stage, the coding disagreements were resolved by discussing the controversial items.

Dependent Variables

The difficulty index b from IRT served as the dependent variable in the current study. Each test item had its own b -value that is typically distributed from -3 to +3 and with a mean of 0 and a standard deviation of 1. A test item with a negative b -value is easier than the one with a positive b -value. For example, an item with a b -value of +2.5 is more difficult than the item with a b -value of -2.5.

Results and Discussion

Table 1 presents the correlation between the difficulty index b and the 17 variables.

Table 1 *Correlation of Variables with The item difficulty (b)*

b	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
b	-0.01	0.08	0.08	0.11	-0.09	0.05	-0.12	0.02	0.12	-0.15	-0.07	-0.03	-0.31	-0.18	-.516**	0.27	-0.30
V1		-0.11	-0.12	-0.02	0.26	-0.10	0.16	0.10	-0.05	0.27	.409*	-0.17	0.32	.476**	0.00	-0.19	-0.04
V2			.377*	0.25	0.11	0.07	0.04	-0.07	-0.10	-0.12	-0.09	0.13	-0.09	-0.12	-0.09	-0.05	-0.09
V3				.672**	-0.05	0.19	0.11	-0.19	-0.14	-0.03	-0.25	0.04	-0.04	-0.33	-0.04	0.13	0.11
V4					0.16	0.10	0.17	0.25	-0.19	-0.01	-0.08	0.14	-0.03	-0.23	-0.19	0.01	-0.08
V5						0.10	.389*	0.08	0.16	0.32	0.04	0.17	0.00	-0.08	0.00	-0.26	-0.11
V6							.603**	-0.30	0.24	0.13	0.03	-0.15	-0.20	-0.22	0.20	0.13	0.22
V7								-0.25	0.15	0.17	0.13	-0.19	0.12	-0.08	0.12	0.08	0.13
V8									0.04	0.17	0.17	0.15	0.00	0.05	0.00	-0.13	-0.03
V9										0.23	0.15	0.05	-0.05	-0.12	-0.17	0.17	0.03
V10											-0.25	-0.02	0.05	-0.16	0.19	-0.21	0.00
V11												-0.29	0.07	.610**	0.24	0.32	.352*
V12													-.355*	-0.32	-0.21	-0.19	-0.29
V13														0.24	0.11	-0.15	0.07
V14															0.24	-0.01	0.18
V15																0.10	.748**
V16																	.563**
V17																	

Note. V1: Idiomatic expressions; V2: Negatives used by the first speaker; V 3: Negatives used by the second speaker; V 4: Negatives in the total dialogue; V 5: Complex sentences used by the first speaker; V 6: Complex sentences used in the second speaker; V 7: Complex sentences used in the total dialogue; V 8: Within-clause referentials; V 9: Between-clause referentials; V 10: Inter-speaker referentials; V 11: Rhetorical questions; V 12: Overlapping words; V 13: Near-synonyms; V 14: Implicit statement; V 15: Propositionally-opposite pair (key-distracter); V 16: Propositionally-opposite pair (distracter-distracter); V 17: Propositionally-opposite pair (any two choices)

* $p < .05$

** $p < .01$

As shown in Table 1, only Variable 15 was significantly and negatively correlated with the item difficulty, suggesting that if the key is propositionally opposite to any of the three distracters, the test item will be easier. The other 16 variables did not significantly correlate with the item difficulty. The result was unsurprising in that the number of the test items was rather small (only 35 items) compared to that in Kostin's (2004) study, where 365 TOEFL items were employed as the materials. The small sample size may weaken the statistical power; therefore, the

disassociation between the 15 variables and the item difficulty was observed in the present study.

Apart from the correlation analysis, the present study also examined how much variance in the item difficulty can be accounted for by the 17 variables. Multiple regression (including the Enter method and the Stepwise method) was adopted to achieve this goal. One caveat should be put forward before multiple regression was conducted; that is, Variable 15 and Variable 17 were highly correlated ($r = .748, p < .01$), which may produce an effect called multicollinearity. Multicollinearity often confounds regression analysis in such a way that any potential individual predictor may become invalid. To avoid multicollinearity, we excluded Variable 17, the one that did not correlate with the item difficulty, from our later analysis. In the Enter method, the difficulty index b -value served as the dependent variable, and the 16 predictor variables were entered as a set into the model. Overall, using the method, no significant model was observed ($F(16, 18) = 1.29, p > .05$, Adjusted R square = .115). The Stepwise method was adopted to identify the minimum number of variables that could be used to predict the dependent variable. In this method, the difficulty index b -value served as the dependent variable, and the 16 variables were entered one by one into the model, with the variable that had the strongest correlation with the dependent variable being entered first. Variables 15 and 16 were left in the final regression equation. The result is shown in Table 2. Together, the two variables accounted for about 33% of the variance with an $F(2, 34) = 9.51, p = .001$.

Table 2 *Results of Stepwise Multiple Regression*

	B	Std. Error	Beta	t	Sig.
Constant	-1.08	.26		-4.18	.00
V15	-2.22	.57	-.55	-3.90	.00
V16	1.90	.81	.33	2.36	.00

Note. Multiple $R = .61$; $R^2 = .37$; Adjusted $R^2 = .33$; standard error of estimate = 1.34.

Although most of the variables did not correlate with and contribute to the item difficulty, two task-processing variables, V15 and V16, *did* have an effect on the item difficulty. Recall that the two variables dealt with how the choice design in terms of opposite propositions may affect test takers' strategies for choosing the answer. The results indicated that test takers would have less difficulty answering the question when the key and any one of the three distracters have opposite propositions (V15), whereas the question with any two distracters having opposite propositions were more difficult to answer (V16). One possible explanation for the two phenomena is that test takers might be aware that the propositionally-opposite pair was very distinct from the

other two options; therefore, they were very likely to confine their attention only to such a pair while ignoring the other two options to a certain degree when choosing the answer. Once the key is in the propositionally-opposite pair, the probability of choosing the correct answer will be increased (V15). On the other hand, if the pair consists of two distracters, the probability of choosing the correct answer will be decreased (V16). This explanation is further supported by test takers' responses to the items coded as V15 and V16. For example, in the case of Variable 15, the distracter from a propositionally-opposite pair (key + distracter) attracted more responses than any of the other two distracters. Consider the item coded as Variable 15 shown below:

男：各位，多虧你們這段時間的努力，終於讓公司今年的營業額增加了，
經理決定多發一個月的年終獎金給大家。

女：他這次這麼看得開，平常一毛不拔的人，怎麼慷慨起來了！

男：從這段對話，可以知道什麼？

(A) 經理是個認真的人 (7) →distracter

(B) 經理是個大方的人 (23)→distracter

(C) 經理是個小氣的人 (72)→key

(D) 經理是個樂觀的人 (3) →distracter

Note. The number in the brackets is the number of test takers who selected that option.

The above example shows that Option B attracted more responses (i.e., 23) than Option A (i.e., 7) or Option D (i.e., 3), revealing that test takers tended to choose the distracter from the propositionally-opposite pair (i.e., Consequence B) more often than from the other two distracters. About 70% of the items coded as Variable 15 displayed such a tendency. Similar tendency was also observed in items coded as Variable 16. A case of Variable 16 is given below:

男：回國以後找工作難不難？

女：一般的工作應該不難，可是，如果想找這樣打從心裡喜歡的工作可不容易。

男：這位小姐的狀況怎麼樣？

(A) 她想要到國外工作 (10) →distracter

(B) 她喜歡現在的工作 (16) →key

(C) 她找工作並不太難 (32) →distracter

(D) 她好不容易找到工作 (46)→distracter

Note. The number in the brackets is the number of test takers who selected that option.

In this example, distracters C and D formed a propositionally-opposite pair, and

attracted more responses than the key and Option A. If test takers' attention had not been directed to the propositionally-opposite pair, it would not have been easy to observe the phenomenon shown in the above example. Therefore, it should be reasonable to conclude that test takers paid more attention to the propositionally-opposite pair than to the other two options. These findings regarding Variables 15 and 16 may provide item writers with a basis for adjusting the items to the desired difficulty level. Specifically, item writers can design a distracter that is semantically opposite to the key if they want to decrease the item difficulty. On the other hand, if they would like to make an item harder, they can produce two distracters with opposite propositions.

This study is among the first studies that attempt to explore the potential factors affecting the item difficulty of short-dialogue comprehension in the Test Of Proficiency-Huayu. Though not as satisfying as expected, the results of the present study at least can provide item writers with some ideas of choice design. However, the observed relation between the item difficulty and item characteristics should be tested in a more empirical manner, such as directly investigating the effect of manipulating the variables identified in the present study on the item difficulty, in order to get a more comprehensive understanding of how those variables may affect the item difficulty. If the results of the present study can be replicated in the future empirical study, we can use them to develop a more practical item-writing guideline for item writers to refer to when creating test items.

References

- Abrahamsen, E., & Shelton, K. (1989). Reading comprehension in adolescence with learning disabilities. *Journal of Learning Disabilities*, 22, 569-572.
- American Heritage Dictionary of the English Language* (4th ed.). (2000). Boston: Houghton Mifflin Co.
- Chaudron, C. (1995). Academic listening. In D. Mendelsohn and J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 74-96). San Diego, CA: Dominie Press, Inc.
- Chiang, C., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*, 26, 345-374.
- Dong, F.-L. (董付蘭). (2000). 談反問句及其教學, *首都師範大學學報, 增刊*, 101-105.
- Kostin, I. (2004). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *TOEFL Research Report RR-04-11*. Princeton, NJ: Educational Testing Service.
- LTTC. (1999). *全民英語能力分及檢定測驗研究：中級測驗預試報告*.
- Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. (TOEFL Research Rep. No. 51). Princeton, NJ: ETS.
- Saeed, J. (2002). *Semantics (2nd edition)*. Oxford: Blackwell.
- Shao, J.-M. (邵敬敏). (2001). *現代漢語通論*. 上海:上海教育出版社.
- Simple and complex sentences. (n.d.). Retrieved July 3, 2008, from <http://users.chariot.net.au/~michaelc/complex.htm>.
- Zhao, L. (趙雷). (2000). 談反問句教學. *語言教學與研究*, 3, 26-31.