The influence of task structure on oral accuracy

Heien-kun Chiang
Feng-lan Kuo
Hui-jun Chen
National Changhua University of Education, Taiwan

This paper reports the findings on the impact of task structure on test takers' oral accuracy. A direction giving task considered as the structured task and a story telling task categorized as the less structured task from the TOEFL *SPEAK* test were used. Five English native speakers and two intact classes of 11th grade Chinese EFL learners participated in this study. To find out the effect of oral proficiency levels on learners' oral performance, only oral production from the native speakers, the higher proficiency learners, and the lower proficiency learners was included in the analysis. Results showed that test takers' oral proficiency interacted with degree of task structure and both factors affected their accuracy and fluency performance. Centering on understanding the impacts of task structure on accuracy components, errors per 100 words and the percentage of error-free clauses, the structured task was found to elicit more accurate speech than the less structured one from the native speakers and the higher proficiency learners. However, the lower proficiency learners produced significantly more accurate utterances in the less structured task than in the structured task. It is likely that these lower proficiency subjects applied avoidance strategies in completing the less structured task.

　本論では、タスクの構成度がテスト受験者のスピーキングの正確さに与える影響を検証する。TOEFLのSPEAKテストの中で、構成度の高いタスクとみなされている道順を伝えるタスクと、より構成度の低いとされるストーリーテリングタスクを使用した。本研究には、5名の英語ネイティヴスピーカーと中国人英語学習者の11年生の2クラスが参加している。学習者のスピーキング習熟レベルが、実際のスピーキングに与える影響を確かめるために、本分析ではネイティヴスピーカー、高レベルの学習者、低レベルの学習者のデータを対象としている。テスト受験者のスピーキング習熟度とタスクの構成度は相互作用し、両方が正確さや流暢さに影響を与えるという結果が出た。タスクの構成度が正確さに与える影響を見ると、構成度の低いタスクよりも、構成度の高いタスクはネイティヴスピーカーや高レベルの学習者から、より正確な発話を引き出すとわかった。しかしながら、低レベルの学習者は、構成度の低いタスクの方が明らかにより正確な発話を産出した。これら低レベルの学習者は、構成度の低いタスクを行うときに回避ストラテジーをたぶん使用するのであろう。
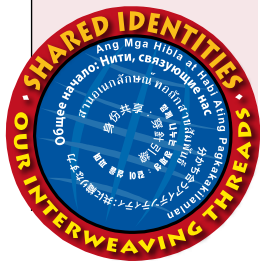
Recently, a volume of research in both ESL and EFL contexts has focused on investigating the impacts of task structure in oral assessments on test takers' oral performance (Chiang, Kuo, Huang, & Chen, 2006; Ejzenberg, 1990, 1997, 2000; Iwashita, McNamara, & Elder, 2001; Kuo, Huang, & Chiang, 2006; Mehnert, 1998; O'Laughlin, 1995; Skehan & Foster, 1999; Teng, 2007). The findings of previous work all indicated that task structure did influence test takers' oral performance to a certain degree in terms of fluency, accuracy, complexity, or lexical density. However, the findings to date as to what aspects of task structure affect test takers' oral performance still remain unclear.

Among the four measurement factors, results of previous studies (Foster & Skehan, 1996; Iwashita et al., 2001; Mehnert, 1998; Skehan & Foster, 1999; Teng, 2007; Wu, 2005) indicated that the effect of task structure on test takers' oral accuracy was rather mixed. Therefore, further investigation on how test takers' oral accuracy was affected by task structures deserves further exploration. To obtain more insight into the relationship between task structure and test takers' oral accuracy, as well as provide empirical evidence to this research area, the examination of how Chinese EFL (henceforth C-EFL) learners' oral accuracy differed in structured and less structured tasks is the first concern in this study.

Additional findings likewise suggest that test takers' oral proficiency seems to interact with the task structure to affect test takers' oral performance (Foster & Skehan, 1996; Iwashita et al., 2001; Mehnert, 1998). Therefore, investigation on the relationship between C-EFL test takers'

oral proficiency and their oral accuracy performance in both the structured task and the less structured one is the second focus of the present study. To examine the relationship among task structure, C-EFL test takers' oral proficiency, and C-EFL test takers' oral accuracy performance, the following two research questions were proposed.

1. Does degree of task structure affect test takers' oral accuracy performance?

2. Is there any relationship between degree of task structure and test takers' oral proficiency levels?

Literature review

In this section, studies related to the impact of task structure on test takers' oral performance are reviewed. It firstly reviews studies concerning the oral accuracy measurements. Two major measures of accuracy, the percentage of error-free clauses and the number of errors per 100 words, will be introduced. Based on the criteria proposed in previous studies, it then categorizes the two oral tasks employed in the present study in terms of degree of task structure. In addition, mixed results of prior studies regarding empirical findings on the effects of task structure on test takers' oral accuracy will be presented.

Oral accuracy measures

Oral accuracy can be measured by focusing on different characteristics of speakers' speech such as the grammatical features of the agreement of noun-modifier, and the usage of the appropriate verb tense in the given context and so on.

Ellis (1987), for example, conducted a study to investigate the speakers' usage of three past-tense forms. Crookes (1989) also conducted a study to examine speakers' usage of the articles and the agreement of third-person singular subject and its modifier. Though specific grammatical points could be the criterion for measuring the speakers' accuracy, they were not suitable measurement features to apply in tasks focusing on communicative purpose. To measure general accuracy, Foster and Skehan (1996) proposed a more reliable and reasonable measure to evaluate oral accuracy under different task conditions. They suggested that accuracy could be measured by the percentage of error-free clauses. Foster and Skehan further defined error-free clause as "clause in which there is no error in syntax, morphology, or word order. Errors in lexis were counted when a word used was incontrovertibly wrong. In cases of appropriateness, no error was recorded" (p310).

In addition to the percentage of error-free clauses as an appropriate measure of accuracy, Mehnert (1998) added a second criterion to evaluate test takers' oral accuracy. That was to count test takers' errors per 100 words. In this study, Mehnert calculated all errors in syntax, morphology, and lexical choice, but not errors in intonation or pronunciation. He further proposed that incorrect inflectional endings be counted as errors, and a repeated error was counted only once. Mehnert's measurement of oral accuracy was adopted by Wu (2005) in a subsequent study.

Thus, in terms of the completeness of the two measurements for evaluating test takers' oral accuracy, Mehnert's (1998) measurement of accuracy was adopted in the present study. In other words, test takers' oral accuracy was measured in terms of the percentage of error-free clauses and the number of errors per 100 words. Errors in syntax, morphology, word order, lexical choice, and inflectional endings were counted, but not the errors in intonation or pronunciation. Furthermore, a repeated error was counted only once.

Degree of task structure

According to Mehnert (1998), Skehan and Foster (1999), and Teng (2007), oral assessment tasks could be classified as structured tasks or less structured ones respectively. To categorize a task as a more structured one or a less structured one, the researcher should consider to what degree a task gives test takers constraints. The present study followed the criteria of Mehnert (1998), Skehan and Foster (1999) and Teng (2007) in categorizing the test tasks utilized in current study as a structured one and a less structured one. Based on the criterion proposed in previous studies, a structured task would provide test takers with more direct cues or predictable storyline in visual stimuli such as pictures, and would supply more constraints for the test takers to complete the task than the less structured one. With direct cues, a predictable storyline and more constraints, test takers' flexibility on completing a task would be reduced. In contrast, a less structured task would supply less direct and less predictable cues in their visual stimuli for test takers to complete the task. Thus, test takers would obtain more flexibility on completing the less structured task. In other words, test takers would have to make decisions on choosing appropriate lexical items or sentence structures or even have to use their imagination to make a meaningful connection

among a set of pictures. On the basis of the afore-mentioned criterion, this study considered a direction giving task based on a pictorial map to be the structured task; and a story telling task based on sequential pictures to be a less structured task.

Effects of task structure on test takers' oral accuracy

Foster and Skehan (1996) adopted three tasks, including a personal information exchange task, a narrative task, and a decision-making task (from structured to less structured) to explore the effects of the degree of task structure on test takers' oral performance in terms of fluency, accuracy, and complexity. In that study, the personal information exchange task was considered as the easiest task followed by the narrative task, and the decision-making task as the most difficult one. Thirty-two pre-intermediate-level EFL students of various L1 backgrounds participated in their study. The result showed that the most structured task – the personal information exchange task – elicited the highest percentage of error-free clauses.

In contrast, Mehnert (1998) utilized an instruction task and an exposition task to investigate the interaction among task familiarity, task structure, and planning time. Subjects of the study were 31 university students studying intermediate level German. Sixteen of them were native speakers of English and the others were of other L1 backgrounds. The instruction task was considered as a more structured and more familiar task, whereas the exposition task was treated as a less structured and less familiar task. The finding was in line with Foster and Skehan's 1996 study in that a structured task elicited more accurate oral output. Mehnert (1998)

suggested that when the task provided test takers with more clear structure and did not ask test takers to express complex ideas, the priority would be given to the fluency or accuracy; that is, test takers would speak more fluently or accurately in completing a more structured task than under a less structured one.

Furthermore, Skehan and Foster (1999) examined the relationship between task structure and task takers' oral performance in terms of fluency, complexity, and accuracy. The subjects were 47 young adults of lower-intermediate English proficiency. They had a variety of first language backgrounds and studied English as a foreign language. In the study, a structured task and a less structured task were utilized. They labeled the task structure degree in accordance with the predictability of the plot of the material eliciting speakers' narration. The results showed that the subjects spoke more fluently in the more structured task. However, task structure did not have a significant influence on speakers' oral accuracy. They claimed that accuracy was likely to be influenced by the interaction of task structure and chances for task preparation.

Building on the findings of Skehan and Foster's research (Foster & Skehan, 1996; Skehan & Foster, 1999), Iwashita et al. (2001) conducted a study to investigate the impacts of task structure on 36 pre-university ESL students' oral performance. In the study, 8 tasks were utilized to stimulate the candidates' speech production. Opposite to the findings of Foster and Skehan's 1996 study, Iwashita et al. found that the less structured task elicited a higher percentage of error-free clauses in their participants' utterance than the structured task did.

In an EFL context, Teng (2007) explored the impacts of task type on 30 Taiwanese EFL college students' oral performance. The subjects studied at the Department of Applied Foreign Languages and had approximately a high-intermediate level of EFL proficiency. A question answering task, a picture description task, and a presentation task were used in her study. The question answering task was considered to be the most structured task since the speakers obtained more detailed and restricted task instructions on how to answer the questions. The results showed that the degree of task structure did not have a significant influence on the test takers' accuracy of performance.

From the aforementioned studies, several issues need to be addressed. First, tasks used in various experiments are not the same, thus the test results may not be directly comparable. Second, the proficiency level and language background of subjects differ. The studies cited above did not utilize an independent test to assess learner's oral proficiency, and they investigated only one level of proficiency. Comparison of oral accuracy between subjects of different ability levels is worthy of examination. Furthermore, some studies examined more than one variable on oral performance. It is likely that the interaction among task structure and the other variables might influence test takers' oral accuracy. Among the studies reviewed above, few were conducted on Chinese subjects and none of them discussed the comparison between levels of proficiency. Therefore, the present study is conducted to find out the effects of task structure on oral accuracy performance of C-EFL learners of varying proficiency levels and to probe the possible interaction between degree of task structure and test takers' oral proficiency levels.

Method

Participants

Participants of this study include C-EFL subjects and native speakers of English. The C-EFL participants are two intact classes of 11th grade students from a senior high school in central Taiwan. *Class A* consists of 39 students, with 19 female students and 20 male students, whereas *Class B* is composed of 35 students, including 17 female students and 18 male students. In total, there were 74 C-EFL participants, consisting of 36 female and 38 male students.

The average Chinese subject is 17 years of age, speaking either Chinese or Taiwanese as their native language and none staying in English-speaking countries longer than one month. These subjects received at least four years of formal English instruction at school, including three years in junior high school, and one and a half years in their senior high school. Since most of the subjects received non-official English instruction in cram schools, the average length of English learning experience for the C-EFL participants is 5.1 years.

To ensure homogeneity, the participants of the two classes were asked to take the *GEPT Intermediate Level Oral Test*, a standardized oral proficiency test developed by *Language Teaching and Testing Center* in Taiwan. The result of the *GEPT Intermediate Level Oral Test* showed that the participants of the two classes are homogenous regarding their oral proficiency level. The participants' scores on the *GEPT Intermediate Level Oral Proficiency* test ranged from 50 to 90 points. The mean score of *Class A* was 65.89, whereas that of *Class B* was 68.00. The result of the *independent t*-test on the scores between the two groups'

Table 1. Results of the T-test on the proficiency score

Class	N	M	SD	df	t
Class A	39	65.89	11.17	72	-.755
Class B	35	68.00	12.78		

Note. p > .05

oral proficiency test was shown in Table 1 (*t* = -.755*, p >* .05). The results thus indicated that there was no significant difference between the two groups, and that *Class A* and *Class B* are at a similar oral proficiency level.

The participants' responses were sent to a native speaker of English for rating. According to the C-EFL participants' scores in the *GEPT Intermediate Level Oral Test*, the C-EFL senior high school learners were further categorized into three groups, including the Chinese EFL High Proficiency learners (henceforth referred to as C-EFL Hi learners), the Chinese EFL Intermediate Proficiency learners, and the Chinese EFL Low Proficiency learners (henceforth referred to as C-EFL Lo learners). The C-EFL Hi learners were the learners who obtained the top 30% *GEPT* scores, and the C-EFL Lo learners were the learners whose *GEPT* scores were at the bottom 30%.

On the other hand, five native speakers of English were invited to take part in the present study. The average age of the native speaker participants was 31 years old. Two of the native speakers were from Canada and the others were from the United States. All of the native speakers teach English in Taiwan, averaging 2.7 years English teaching experience in Taiwan. The speech samples of the native speakers of

English served as the baseline for comparison with the C-EFL participants' speech outputs.

Instruments

The instruments utilized in this study include a *GEPT Intermediate Level Oral Test*, and the TOEFL *SPEAK* (*Speaking Proficiency English Assessment Kit*) subset test, a standardized oral proficiency test launched by *Education Testing Center*. The *GEPT Intermediate Level Oral Test* was used to ensure the homogeneity between the two classes of subjects. In contrast, the *SPEAK* test can be categorized into three parts. The first section is a pictorial map description task in which the test takers were asked to describe the information which the pictorial map provides. The second section is a story telling task based on sequential pictures. In the story telling task, the test takers were provided with a series of six pictures, and they were asked to tell the story in accordance with the pictures and then were asked some follow-up questions. The third part is a statistical chart or graph reading task. In this task, the test takers were given a statistical chart or graph and asked to describe or explain the information embedded in it, then answer follow-up chart or graph-based questions. However, in the current study, the statistical chart or graph description task was excluded since previous study indicated that a person's ability to describe embedded chart or graph related information requires training (Xi, 2005). Thus, only the pictorial map description task and the sequential picture description task were performed by the participants.

According to the pattern of studies such as Mehnert (1998), Skehan and Foster (1999), and Teng (2007), the

direction giving task was considered as the structured task in the present study. In the direction giving task, the test takers obtain a lot of cues from the pictorial map such as the names of the streets and the names of the buildings in the small town, etc. Furthermore, for this task, the test takers are required to respond to the question directly according to the information provided in the map so that they have less flexibility for their utterances. On the other hand, the story telling task is considered as a less structured task since the test takers have more flexibility to make up their stories and they have greater freedom for their narrations. Though the sequential pictures provide the test takers with the storyline, the test takers still have the flexibility with their narrations and freedom with their lexical choices. Therefore, the story telling task in the present study was considered as a less structured task (Mehnert, 1998; Skehan & Foster, 1999; Teng, 2007).

Data collection procedure

Initially, the participants completed a personal background information questionnaire (see Appendix 1). To complete the questionnaire, the participants were given fifteen minutes to respond to the thirteen questionnaire items. Then the participants completed the *GEPT Intermediate Level Oral Test* which was followed by the TOEFL *SPEAK subtest*. When taking the *GEPT Intermediate Level Oral Test* and the *SPEAK subtest*, the participants heard all of the questions played aloud over classroom loudspeakers and all of the participants' oral responses were recorded by a computer program, *Ulead Video*.

Data analysis

The researchers collected and analyzed the speech samples of the native speakers of English, the C-EFL Hi learners, and the C-EFL Lo learners in the direction giving task based on the pictorial map and the speech samples in the story telling task based on the sequential pictures in the *SPEAK subset*. The speech samples were transcribed and coded by the third author for oral accuracy analysis. Participants' oral accuracy was measured in terms of the number of errors per 100 words, and the percentage of error-free clauses (Foster & Skehan, 1996; Mehnert, 1998). For statistical analyses, two-way ANOVA was used to investigate the influence of task structure and the possible interaction between task structure and subjects' oral proficiency levels on the participants' oral accuracy performance. When the results showed significant differences for the proficiency factor, post hoc analyses followed in order to further examine the differences among the oral accuracy performance of the native speakers of English, the C-EFL Hi learners, and the C-EFL Lo learners in the two tasks.

Results and discussion

The results are first presented in terms of the number of errors per 100 words produced by the NS, the C-EFL Hi learners, and the C-EFL Lo learners. Then the oral performances of the three groups are compared according to the percentage of error-free clauses they produced.

Errors per 100 words

Table 2 shows the means and standard deviations of the number of errors per 100 words produced by the NS, the C-EFL Hi learners, and the C-EFL Lo learners in the direction giving task and in the story telling task. As shown in Table 3, the ANOVA analysis did not yield a significant difference between the number of errors per 100 words found in the two tasks ($F = .66$, $p > .05$). On the other hand, the results showed a significant difference among the three proficiency groups ($F = 75.14$, $p < .05$). As found in Table 2, the mean scores of the NS and the C-EFL Hi groups were lower than those of the C-EFL Lo in both the direction giving task and the story telling task. This indicates that subjects' oral proficiency level was a major factor affecting their oral accuracy. To further examine the differences among the three proficiency groups, post hoc analyses were conducted and they showed significantly fewer errors per 100 words for the NS and the C-EFL Hi groups. The oral accuracy analysis also disclosed a significant interaction effect between the degree of task structure and subjects' oral proficiency levels, suggesting that the number of errors per 100 words differed between the higher proficiency groups and the lower proficiency groups. The NS and the C-EFL Hi groups indeed produced significantly less errors per 100 words in the direction giving task. In contrast, the C-EFL Lo learners produced significantly less errors per 100 words in the story telling task. The results indicate that the structured task (the direction giving task) elicited significantly more errors from the C-EFL Lo learners than the less structured task (the story telling task) did. The results in fact deviated from the findings of the previous studies (Foster & Skehan,

1996; Mehnert, 1998). Previous studies showed that the structured task would elicit more accurate utterances from the test takers than the less structured task did. However, Riazantseva (2001) suggested that test takers might apply avoidance strategy in completing the less structured task due to fewer constraints were given in the less structured task. In the current study, the C-EFL Lo learners seem to apply the avoidance strategy when completing the story telling task (less structured task).

Table 2. Means and standard deviations of the number of errors per 100 words for NS, C-EFL Hi learner, C-EFL Lo learner groups in the direction giving task and the story telling task

Language group	n	Direction giving		Story telling	
		M	SD	M	SD
NS	5	0.95	0.67	2.14	0.46
C-EFL Hi learners	22	4.19	3.15	5.43	1.84
C-EFL Lo learners	22	23.70	11.20	16.58	5.65

Table 3. Two-way ANOVA summary for task structure and proficiency in errors per 100 words measure

Source	DF	SS	MS	F	P
Task	1	26.00	26.00	.66	.42
Proficiency	2	5887.18	2943.59	75.14	.00*
Task x Proficiency	2	371.98	185.99	4.75	.01*

The percentage of error-free clauses

Table 4 illustrates the descriptive statistics of the percentage of error-free clauses produced by the NS, the C-EFL Hi learners, and the C-EFL Lo learners in the direction giving task and in the story telling task. As shown in Table 4, the ANOVA analysis revealed a significant difference between the percentage of error-free clauses found in the two tasks (F = 6.07, $p < .05$). The percentage of error-free clauses found in the direction giving task was significantly higher than that found in the story telling task, indicating a more structured task elicited less errors. Additionally, the results showed a significant difference among the three proficiency groups ($F = 84.31$, $p < .05$). As seen in Table 4, the mean scores of the NS and the C-EFL Hi groups were higher than those of the C-EFL Lo in both the direction giving task and the story telling task. This indicates that subjects' oral proficiency level was a major factor affecting their oral accuracy. To further examine the differences among the three proficiency groups, Scheffe multiple group comparisons were conducted and they showed a significantly higher percentage of error-free clauses for the NS and the C-EFL Hi groups. The oral accuracy analysis also displayed a significant interaction effect between degree of task structure and subjects' oral proficiency levels, indicating that the percentage of error-free clauses differed between the higher proficiency groups and the lower proficiency groups across the two different tasks. The NS and the C-EFL Hi groups indeed produced significantly higher percentages of error-free clauses in the direction giving task, however, the two groups did not differ in the percentage of error-free clauses they produced within each task. In contrast, the C-EFL Lo learners produced an

Table 4. Means and standard deviations of the percentage of error-free clauses for NS, C-EFL Hi learner, C-EFL Lo learner groups in the direction giving task and the story telling task

Language group	n	Direction giving		Story telling	
		M	SD	M	SD
NS	5	89.13%	10.58%	78.01%	4.62%
C-EFL Hi learners	22	78.61%	16.54%	57.68%	11.48%
C-EFL Lo learners	22	22.69%	21.00%	20.84%	22.05%

Table 5. Two-way ANOVA summary for task structure and proficiency in the percentage of error-free clauses measure

Source	DF	SS	MS	F	P
Task	1	1995.27	1995.27	6.07	.02*
Proficiency	2	55415.94	27707.97	84.31	.00*
Task x Proficiency	2	2355.01	1177.50	3.58	.03*

Table 6. Comparison of the oral accuracy display of NS, C-EFL Hi learners, and C-EFL Lo learners

Measurement feature	Direction giving	Story telling
Errors per 100 Words	NS=Hi<Lo	NS=Hi<Lo
Percentage of error-free clauses	NS=Hi>Lo	NS=Hi>Lo

approximately equal percentage of error-free clauses in the two tasks. This suggests that the lower proficiency subjects were not able to take advantage of the cues provided in the more structured task.

Table 6 summarizes the findings of the interaction between test takers' oral proficiency and the influence of task structure on test takers' oral accuracy. Table 6 showed that the C-EFL Hi learners produced similar number of errors with the NS group in both tasks. However, the C-EFL Lo learners produced significantly more errors per 100 words than the other two groups. Regarding the percentage of error-free clauses produced by the three groups, the NS group and the C-EFL Hi learners produced a significantly higher percentage of error-free clauses than the C-EFL Lo learners in both tasks.

Conclusion and future study

The results of the current study showed that structured task elicited more accurate utterances from the NS group and the C-EFL Hi learners than the less structured one. The C-EFL Lo learners, however, are more likely to apply avoidance strategies in the less structured task; thus, they produced more accurate speech in the less structured task than in the structured one.

According to the findings of the current study, there were some pedagogical implications for the English oral assessment development and English teaching and learning in EFL environments. First of all, the results of the current study revealed that degree of task structure (structured vs. less structured) had an impact on test takers' oral accuracy; therefore, it can be suggested that the English oral assessment developers who take accuracy as the first priority in measuring

take task structure into consideration when developing new oral assessments. Furthermore, it is suggested that EFL instructors familiarize their learners with different kinds of tasks in their oral training courses, which could reduce difficulties raised by test takers' unfamiliarity to task types.

However, there are some limitations of the present study. First, the sample size could be enlarged. Due to the limitation of human energy and time, this study only invited two classes of 11th graders in Taiwan as the C-EFL learner samples and 5 native speakers as the comparison. In future studies, researchers could involve larger number of participants to obtain more general speech patterns. Second, more tasks could be included in future studies. Since the present study has shown that task structure had a great influence on test takers' oral performance, further studies could explore the impacts of other types of tasks on test takers' oral performance. Finally, interviews or questionnaire instruments exploring participants' perceptions toward the influence of task structure are also worthy of investigation.

Heien-kun Chiang completed his Ph.D. degree in the Computer Science Department at the University of Illinois at Urbana-Champaign in 1997. His research interests include internet technology, e-learning, and intelligent tutoring and diagnosis systems. He is currently an associate professor in the Information Management Department at National Changhua University of Education, Taiwan. <hkchiang@cc.ncue.edu.tw>

Feng-lan Kuo received her Ph.D. degree in the Linguistics Department at the University of Illinois at Urbana-Champaign in 1994. Her research interests include

phonetics, phonology, phonics, and web-based English teaching and learning. She is currently serving as a professor in the English Department at National Changhua University of Education, Taiwan. <laflkuo@cc.ncue.edu.tw>

Hui-jun Chen received her MA degree in TESOL at National Changhua University of Education, Taiwan in 2007. She is currently serving as a lecturer in Ling Tung University, Taiwan. <violetbuglet@yahoo.com.tw>

References

Chiang, H. K., Kuo, F. L., Huang, M. Y., & Chen, H. J. (2006). A comparative study on lexical richness and lexical density in two spontaneous speeches of Chinese EFL learners. *SPECTRUM: NCUE Studies in Language, Literature, Translation, and Interpretation, 1*, 55-71.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.

Ejzenberg, R. (1990). Oral fluency in the context of second/foreign language testing: Understanding the role of task structure and formulaic speech. (ERIC Document Reproduction Service No. ED334238)

Ejzenberg, R. (1997). The role of task structure in oral fluency assessment. (ERIC Document Reproduction Service No. ED406827)

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287-313). Ann Arbor, MI: University of Michigan Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, *9*, 12-20.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51*, 401-436.

Kuo, F. L., Huang, M. Y., & Chiang, H. K. (2006). Task-related variation in interlanguage: The case of pauses in speech production. *Proceeding of the 23rd International Conference on English Teaching and Learning in the Republic of China*, 637-652, Taipei, Taiwan: Crane.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 83-108.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing*, *12*, 217-237.

Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition, 23*, 497-526.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93-120.

Teng, H. (2007). *A study of task types for L2 speaking assessment*. (ERIC Document Reproduction Service No. ED496075)

Wu, R. (2005). Effect of familiarity on spoken performance in a picture description task. *Proceedings of the 22nd International Conference on English Teaching and Learning in the Republic of China*, 330-341, Taipei, Taiwan: Crane.

Xi, X. (2005). Do visual chunks and planning impact performance on the graph description task in the SPEAK exam? *Language Testing, 22*, 463-508.

Appendix 1

Language Background Questionnaire

Dear participants,

We would like to understand your English learning experiences. Please help us by answering the following questions. Thanks for your time!

Hui-chun Chen

Graduate Student of Department of English

National Changhua University of Education, Taiwan

1. Student number: _____

2. Gender: ☐ male ☐ female

3. Age: _____ years old

4. Native language (the language you learn from birth):

   ☐ Taiwanese Southern Min

   ☐ Mandarin Chinese

   ☐ Hakka

   ☐ Other(s): _____

5. Have you ever lived in an English-speaking country? If yes, how long have you been there?

   ☐ No

   ☐ America _____ years   _____ months

   ☐ England _____ years   _____ months

   ☐ Australia _____ years   _____ months

   ☐ New Zealand _____ years   _____ months

   ☐ Other(s): _____ _____ years _____ months

6. At what age did you start learning English? Age: _____

7. How many years have you learned English until now? Years: _____

8. Where did you first learn English?

   ☐ in a cram school

   ☐ in a public school

   ☐ at home, from parents / siblings

   ☐ at home, from a tutor

   ☐ in a cram school and a public school

   ☐ Other(s): _____

9. If you have ever learned English in a cram school, did you learn from

   ☐ a non-native teacher

   ☐ a native teacher

   ☐ a non-native teacher and a native teacher

10. Have you ever taken General English Proficiency Test (GEPT)?

☐ Yes (Please go to No. 11 and 12)

☐ No (Please go to No. 13)

11. Which level test have you ever taken?

☐ Elementary Level Test, first stage

☐ Elementary Level Test, second stage

☐ Intermediate Level Test, first stage

☐ Intermediate Level Test, second stage

☐ High Intermediate Level Test, first stage

☐ High Intermediate Level Test, second stage

☐ Advanced Level Test, first stage

☐ Advanced Level Test, second stage

12. Which level test have you ever passed?

☐ Elementary Level Test, first stage

☐ Elementary Level Test, second stage

☐ Intermediate Level Test, first stage

☐ Intermediate Level Test, second stage

☐ High Intermediate Level Test, first stage

☐ High Intermediate Level Test, second stage

☐ Advanced Level Test, first stage

☐ Advanced Level Test, second stage

13. Have you ever taken a practice GEPT?

☐ Yes

☐ No