

A Quantitative Study on Effective Speaking Frequency and Hesitation in the Second Language Oral Test Using Man-machine Conversation Mode

Yong Fang^{1,2*}

¹School of Foreign Languages, Jiangnan University

²School of Chinese Language and Literature, Central China Normal University

*Yong Fang. Email: 875540144@qq.com

ABSTRACT

Effective speaking frequency and hesitation of the examinee in the second language oral test are important aspects of their performance. In this study, two groups of undergraduates, five in each group, were recruited as subjects, and six teachers were recruited as examiners. A simulated oral English test was organized based on man-machine conversation mode. Based on the statistics and analysis of the experimental data, it is found that, 1. The distribution of the scores given by the examiner is confirmed by the effective speaking frequency and hesitation duration; 2. In addition to the oral English ability of the subjects, their performance is also affected by the topic. Better familiarity with the topic result in better performance; 3. The effect of familiarity on advanced English learners is significantly higher than that on intermediate English learners.

Keywords: *man-machine conversation mode, second language oral test, effective speaking frequency, hesitation*

1. INTRODUCTION

The economic advantage of man-machine conversation mode is very obvious. With a relatively small number of invigilators, a large number of candidates can be examined. In addition, the examiner does not need to go to the test site in person, which improves the scoring efficiency and reduces the cost. Because examiners and examinee don't meet, it can maximize the guarantee that the score is not affected by other factors, and improve the fairness of the scoring. Science and technology has not yet advanced to the stage of natural discourse communication between computer and human. Researchers have doubts about the man-machine conversation mode, and believe that the biggest defect lies in its non-authenticity. Some researchers have pointed out that in the real world, it is rare for speaker to make a complete speech on the specified topic within the specified time, and candidate's oral performance are very likely to be affected because they are not adapted to this unnatural state of oral communication [1],[2],[3]. Moreover, the result of the evaluation is not equal to the examinee's oral ability in

the natural state because the test is based on unnatural state.

The study on how the candidates' oral performance is affected by variables such as the characteristics of man-machine conversation mode and the examiner's judgment is an important basis to improve the test. The subjects' effective speaking frequency and hesitation are important aspects of their performance in such test. Therefore, Yong Fang [4] designed and implemented a simulated oral English test based on man-machine conversation mode to investigate those aspects.

2. EXPLANATION OF THE EXPERIMENT

Five English majors and five non English majors undergraduates who had the experience of oral English test, and were at the middle level, were recruited as subjects. The purpose of setting the range of the subject's oral proficiency is to avoid the ceiling effect caused by high oral ability, and the failure of detecting caused by low oral ability. The experiment also recruited 6 teachers, including 3 native speakers of Chinese and 3 native speakers of English, as examiners to score the recorded materials. The experiment was conducted in a

relatively closed and quiet environment. After the debugging of the experimental equipment, the subjects made impromptu English speeches according to the two topics given on the screen. Each topic has one minute for preparation and two minutes to deliver. There is a one minute interval between the two topics to prevent experimental fatigue. In order to prevent the possible influence of the order of the two topics on the oral performance of the subjects, this experiment adopts the method of counterbalance test to collect data. All speeches were recorded as research data. After the oral test, the subjects completed a questionnaire concerning the test. The data analysis concerning effective speech frequency and hesitation in this experiment are disclosed as follows.

3. ANALYSIS AND DISCUSSION ON EFFECTIVE SPEECH FREQUENCY OF SUBJECTS

Effective information and fluency are two important assessment points in oral test. At the data analysis stage, the subjects' live recordings are dictated to form a complete text, and then the data are extracted from the text for quantitative analysis.

3.1. Descriptive Statistics of Effective Speech Frequency

Firstly, the number of effective words of the subjects was counted. Effective words refer to lexicon with semantic meaning. Filter words, such as "uh, EM", are excluded in statistics. Because the number of effective words and speech time of each subject are different, there is no statistical significance to compare the absolute number of effective words or the length of speech. In order to make the data comparable, the number of effective words is divided by the speaking time, and the production per second is calculated, which is defined as the effective speaking frequency. The reason for calculating the effective speech frequency is that it can reflect the fluency of the speaker, and fluency is the direct embodiment of the speaker's oral ability. The view was supported by many researchers. For example, Lennon [5] defined the speaker's fluency as the speed of speaking. After statistical analysis of the calculated effective speech frequency, the average effective speech frequency of the ten subjects in two topics is 1.08 words / second (n=10, standard deviation is 0.28 words / second) and 0.87 words / second (n=10, standard deviation is 0.26 words / second) respectively. Table 1 shows the relevant test data.

Table 1. Descriptive Statistics of Effective Speech Frequency

	Average Value	Number (N)	Standard Deviation	Standard Error
Topic 1	1.0810	10	.28290	.08946
Topic 2	.8700	10	.26077	.08246

3.2. Test on Intra Group Difference Value and Correlation

The average speaking frequency of the two topics is different. In order to verify whether the mean difference is statistically significant, an intra group difference test

(independent T-test) was conducted. The results show that the effective speaking frequency of the same subject is significantly different when completing two topics. The Pearson value is $p = .014$, and the reliability value is $r = 0.71$, indicating significant reliability. Table 2 shows the test results.

Table 2. Intra Group Difference Test--- Effective Speech Frequency

	Intra Group Differences					t	df	Significance (two way)
	Average Value	Standard Deviation	Standard Error	95% Confidence Interval of the Difference				
				Low Range	High Range			
Topic1-Topic2 (effective speaking frequency)	.21100	.21896	.06924	.05437	.36763	3.047	9	.014

The phenomenon that the performance of the subjects in two topics is significantly different suggests two possibilities. First, the subjects' performance is inconsistent due to the influence of participant fatigue. Second, the different nature and content of the selected topics lead to difference in difficulty, which leads to differences of subjects' performance. The experiment adopts the method of counterbalance study. If the experimental fatigue has such a significant impact on

the subjects, half of the subjects should show better effective speaking frequency in topic 1 than in topic 2, while the other half should be on the contrary. However, the statistical data as a whole shows no significant difference, so the first possibility can be ruled out. Therefore, the second possibility is further analyzed and discussed as follow.

In order to demonstrate whether there are differences

in difficulty of the test topics, which result in significant differences in the performance of the subjects, the correlation statistical analysis was carried out. If it is confirmed that there is a significant intra group correlation in the effective speaking frequency of the two topics, it shows that the subjects' performance of

the two topics is stable, and the difference is not caused by personal factors, but affected by external variable, the difficulty of the topics. The data of intra group significance test confirms this hypothesis, and the significance value was $p = .031$. Table 3 shows the statistical results.

Table 3. intra group correlation test

	Number	Relevance	Significance
Topic1-Topic2	10	.678	.031

The author believes that the reason for the difference in difficulty between the two topics is determined by the nature of the topics themselves. Specifically, topic 1 is 'the most important decision made by an individual', topic 2 is 'the advantages and disadvantages of television'. The former focuses on personal experience, while the latter is a general comment topic. The former is closer to the life of the subjects, while the latter is relatively far away from the subjects' daily thinking. The subjects' familiarity with topic 1 and their desire for cognitive expression were significantly better and stronger than those of topic 2, resulting in the phenomenon that the subjects performed better in topic 1. It can be further inferred that in order for candidates to better demonstrate their real oral second language ability, the topics should be test taker oriented, close to the candidates' personal life experience or their daily concerns. Only in this way, can we effectively eliminate the phenomenon that the examinees are not familiar with the topic or lacking of desire to talk about it, which leads to the distortion of the examinee's second language oral performance.

3.3. Analysis of Inter Group Oral Performance Differences

Another matter need to demonstrate is the relationship between the subjects' oral ability and their oral performance in the man-machine conversation test. Previous analysis has proved that the scores of advanced English learners are significantly higher than those of intermediate English learners. From the perspective of the effective speech frequency of the subjects, if the analysis of the difference value of the effective speech frequency is also significant, the following points can be concluded directly or indirectly. 1) the man-machine conversation oral can effectively reflect the differences of the examinees' oral ability; 2) the examiners' scores in this experiment are reliable; 3) the effective speech frequency of the speaker can effectively reflect his/her oral ability.

First of all, the difference degree of the average effective speaking frequency of the two topics was statistically analyzed. The results (Table 4) show significant differences between the groups ($p = .045$, $p < 0.05$). The statistical results are highly reliable ($r = .73$).

Table 4. Analysis of Intra Group Oral Performance Differences between Advanced English Learners and Intermediate English Learners

	Mean Difference <i>T-test</i>						
	t	df	Significance (bidirectional)	Average Difference	Standard Deviation Difference	95% Confidence Interval of the Difference	
						Lower	Upper
Average Speaking Frequency	2.552	5.754	.045	.31800	.12462	.00988	.62612

The results in Table 4 are in line with the expectation and confirm the above three sub-arguments. However, the above analysis shows that the correlation between the topic and the subjects' personal life also affects their desire for expression and oral performance. If the analysis is only based on the average speaking

frequency of the subjects, it can't reflect the influence of the test topic differences on the subjects with different oral abilities. Therefore, the differences of the speaking frequency of the subjects in each test topic were analyzed separately, and the results are shown in Table 5.

Table 5. Analysis of Oral Performance Differences between Advanced English Learners Group and Intermediate English Learners Group on Topic 1 and Topic 2

Effective Speak Frequency	Mean Difference <i>T-test</i>						
	t	df	Significance (bidirectional)	Average Difference	Standard Deviation Difference	95% Confidence Interval of the Difference	
						Lower	Upper
Topic 1	2.293	8	.050	.33800	.14742	-.00196	.67796
Topic 2	2.112	4.570	.094	.29600	.14017	-.07475	.66675

Table 5 shows that, from the perspective of effective speaking frequency, there are significant differences between advanced English learners and intermediate English learners in the oral performance of topic 1, $p = 0.05$, and the reliability is moderate ($r = 0.63$). However, there is no significant difference between the two in topic 2, Pearson value is $p = .094$, reliability is significant ($r = 0.70$). Based on the results of Table 2 and Table 4, it can be concluded that advanced English learners are more likely to be affected by the factors concerning non-oral ability, such as their familiarity with the topics, the correlation between the topics and themselves, and the their desire to express the topics.

The above analysis and discussion lead to the following reflection. In the man-machine conversation test, candidates are unable to choose the test topics, lacking of real oral interaction objects, and show their oral ability only by personal statements on given topics. Candidate’s unfamiliarity with the topic and lacking of personal experience, limit their oral performance in the test. Candidates with higher oral ability are more likely to be affected by the above potential non-competence factors. In other words, in the man-machine conversation test, the higher the oral proficiency of the examinee, the more unlikely the candidate’s oral competence can be detected. In view of this, the choice of topics in man-machine conversation mode becomes very critical. The test organizers should investigate the background of the subjects, understand areas that the candidates are familiar with and keen to discuss, and take this as the basis for preparing the test topics. Some researchers believe that topics with abstract scope, which involves no professional fields, can detect candidates' oral competence fairly and effectively. This view is right from the perspective of universality, and it is not contradictory to the suggestions of this study.

4. ANALYSIS AND DISCUSSION OF SUBJECTS' HESITATION

Another dimension of the participants' oral fluency is hesitation. Hesitation includes the complete silence between two expressions, and the time taken by filter word with no semantic meaning in the oral expression process. Tavakoli [6] pointed out that statements without hesitation can be considered as fluent oral expression. According to the above analysis, the more familiar the subjects are with the topics, and the closer the topics are to their lives, the higher their effective speaking frequency is, and the more words they produce per unit time. From the perspective of hesitation, the interpretation of the above is the result of the lower percentage of hesitation duration per unit time.

4.1. Statistics of Subjects' Hesitation

In order to confirm this interpretation, we counted the hesitation of the subjects in two groups. 0.500 second is taken as the dividing line of hesitation in statistics, that is, the silence time over 0.500 second is defined as hesitation. The basis of choosing 0.500 second is that silence below 0.500 seconds may be natural breath or semantic hesitation of the speaker. In addition to that, the hesitation of less than 0.500 second is almost imperceptible, which does not affect smooth communication. Any event with a transitional word is considered as a hesitation, and its duration is calculated from the offset of the previous word of the transitional word to the onset of the next word of the transitional word. According to the above standard, the hesitation of the subjects in topic 1 and topic 2 is counted respectively, and then the average of all hesitations of subjects in a single topic was processed, so as to facilitate the intra and inter group comparison. Table 6 shows the statistical results of hesitation duration after average value processing.

Table 6. Hesitation of Subjects (unit: second)

Subject	Language Competence	Topic 1	Topic 2
1	Advanced English learner	1.229	2.023
2	Advanced English learner	1.258	1.154
5	Advanced English learner	1.209	2.282
6	Advanced English learner	1.187	1.791
10	Advanced English learner	1.118	1.829
3	Intermediate English learner	1.283	1.892

4	Intermediate English learner	1.345	1.767
7	Intermediate English learner	1.477	2.087
8	Intermediate English learner	1.385	2.117
9	Intermediate English learner	1.506	1.881

From the data shown in Table 6, the average hesitation duration of the subjects in topic 1 is lower than that in topic 2. The average hesitation duration of advanced English learners is lower than that of intermediate English learners. In order to verify the statistical significance of

the data, the differences between groups and within groups are analyzed. Table 7 shows the descriptive statistics about all subjects' hesitation in topic 1 and topic 2.

Table 7. Descriptive Data of Subjects' Hesitation (unit: second)

Hesitation	Average Value	Number	Standard Deviation	Standard Error
Topic 1	1.2997	10	.12661	.04004
Topic 2	1.8823	10	.30396	.09612

The descriptive data in Table 7 shows that the hesitation duration of all subjects in topic 1 (average = 1.230 seconds) is significantly shorter than that in topic 2 (average = 1.882 seconds).

4.2. Intra group and Inter group Difference Test of Subjects' Hesitation

The intra group difference test (paired sample test) was performed later, and the statistical analysis results are shown in Table 8. It can be observed that there is a significant difference in the hesitation duration of all subjects in topic 1 and topic 2. The Pearson value is $p < 0.01$, and with a very high reliability ($r = 0.89$). This

shows that, from the perspective of hesitation, the fluency of the subjects in topic 1 is significantly higher than that in topic 2. As pointed out above, subjects are more familiar with topic 1 than topic 2. Therefore, it is speculated that the subjects have more to say about topic 1. In the process of presentation, they need less time to consider the content to present, and therefore can spend more energy on the organization of English vocabulary and sentence structures. The above factors lead to the shortening of hesitation duration of the subjects. The results of data analysis are consistent with the results of effective speech frequency statistics in the previous part. The closer the topic is to the their life, the more familiar the examinee is with the topic, the more outstanding the examinee's oral performance is in the oral test.

Table 8. Intra-group Difference Test of Hesitation

Hesitation	Intra-group Difference				t	df	Significance (bidirectional)	
	Average Value	Standard Deviation	Standard Error	95% Confidence Interval of the Difference				
				Lower				Upper
Topic1-Topic2	-.5826	.31023	.09810	-.80453	-.36067	-5.939	9	.000

The results of intra group differences of hesitation duration were consistent with those of intra group differences of speech frequency. Based on this, we will verify whether the differences between groups have a corresponding relationship with each other in these two

variables. The average hesitation duration and English ability of the subjects are taken as two variables, with the former as the dependent variable and the latter as the independent variable. Table 9 shows the analysis results of the differences between groups.

Table 9. Inter Group Difference Test of Hesitation

Hesitation	T-test of Mean Value between Groups						
	t	df	Significance (bidirectional)	Average Difference	Standard Error	95% Confidence Interval of the Difference	
						Lower	Upper
Topic 1	-4.182	8	.003	-.19900	.04758	-.30872	-.08928
Topic 2	-.670	8	.522	-.13300	.19841	-.59053	.32453

Similar to the results of the inter group difference

test of effective speaking frequency, table 9 shows that

there are significant inter group differences in the average hesitation duration of advanced English learners and intermediate English learners in topic 1, with Pearson value of $p = .003$. The difference has a very high reliability ($r = 0.83$). However, there was no significant difference between the two groups in topic 2 ($P = .522$). Based on the above data and the results that the average hesitation duration of the two groups is significantly different in topic 1 and topic 2, it can be inferred that the subject's familiarity with topics has a more significant effect on Advanced English learners than on Intermediate English learners. The above inference is consistent with the inference about the interaction between the test topics and the candidates' English ability in the previous part. It further demonstrates that the familiarity of the test topics as a non-ability factor is related to the candidates' oral fluency during the test. The familiarity with the topic promote the examinee's performance in the test.

5. CONCLUSION

Throughout the analysis and discussion on the specific oral performance quality of effective speaking frequency and hesitation under the man-machine conversation mode, the conclusions are made as follows. Firstly, the fluency of the subjects reflected by the data, including the effective speaking frequency and the hesitation duration per unit time, confirms the distribution of the examiners' scores on the subjects in the previous study. In other words, advanced English learners are better than intermediate English learners in oral fluency. Secondly, in addition to the oral English ability of the subjects, their performance is also affected by the given topic. The more familiar the subjects are with the topic, the better their oral performance is. Thirdly, familiarity with the content has a greater impact on the oral fluency of advanced English learners than that on intermediate English learners.

REFERENCES

- [1] Shi, F. Exploring Students' Anxiety in Computer-based Oral English Test. *Journal of Language Teaching and Research*, 3(3), 446-451, 2012.
- [2] Qian, D. D. Comparing direct and semi-direct modes for speaking assessment: affective effects on test takers. *Language Assessment Quarterly*, 6(2), 113-125, 2009.
- [3] Fischer, Donald C. Jr. Comparing face-to-face and distance modalities, in: *Conducting Arabic and Russian speaking proficiency tests*. Dissertation Abstracts International, A: The Humanities and Social Sciences, 2005, 4536-A.
- [4] Yong, Fang. An Experimental study on the effectiveness of oral tests for Chinese learners of English, in: *Doctoral Dissertation*, Central China Normal University, May 2015.
- [5] Lennon, P. *The Lexical Element in Spoken Second Language Fluency*. H. Riggenbach (ed.), 2000.
- [6] Tavakoli, P. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, 65/1, 71-79, 2011.