

Linguistic, Grammatical and Lexical Features of Examinees' Responses during Semi-direct and  
Direct Speaking Assessments

Shi Chen

Northern Arizona University

### Abstract

In the current study, the researcher compared a direct speaking assessment and a semi-direct speaking assessment from linguistic, grammatical, and lexical perspectives. Previous research has focused on the score analysis, and validity argument of the two testing delivery modes; however, no research involved analysis of the speaking responses elicited from the two types of delivery modes via acoustic measures. 13 participants were recruited from the program of intensive English (PIE) at Northern Arizona University. The instruments used in this research were the retired TOEFL speaking tests. Each participant took the test face to face with the interlocutor, and took the test through a computer program. 26 speech files were produced in total. The researcher adopted ten acoustic measures, analyzing the speech via pitch, stress, and tone choices. The speech files were also analyzed from grammatical perspectives using global accuracy and grammar complexity. In addition, the researcher also used the Vocabprofile website to analyze the token ratio of the speech files. The results show that most variables do not have much difference except for overall pitch range and coverage of the first one thousand words from the General Service List (GSL).

*Keywords:* direct speaking assessment, semi-direct speaking assessment, speech perception, fluency, tone choice, stress, grammatical features, acoustic measures, lexical measures

Linguistic, Grammatical, and Lexical Features of Examinees' Responses during Semi-direct and  
Direct Speaking Assessments

### **Background**

Direct and semi-direct methods are often adopted to elicit spoken English responses from examinees. The direct method involves a face-to-face interview, and the semi-direct method involves a machine-mediated speaking-assessment instrument. A substantial number of studies have investigated the differences between semi-direct and direct speaking assessments from the perspectives of validity. Quite a few studies have declared that the two modes are equivalent or at least strongly correlated. One goal of the current research is to provide more solid evidence to confirm or reject the equivalence of the two test modes by expanding the measures to linguistic variables. The present study extracted two TOEFL independent speaking tasks and administered them to participants both face-to-face and with a computer.

No prior research appears to have used linguistic measures (the perspective of speech perception) to directly investigate the differences between these two delivery modes. The current research fills the gap by measuring linguistic features, such as fluency, stress, and tone choices. Previous studies have employed the IELTS, the Oral Proficiency Interview (OPI), and self-developed instruments, yet none of them have attempted to use the high-stakes TOEFL speaking tasks as instruments. The results of such an investigation would benefit stakeholders and administrators by determining which type of speaking assessment elicits the most desirable performance and which best suits their administrators: it would also determine whether the two

test modes elicit equivalent linguistic and grammatical features from the test takers. This paper will also discuss the implications that linguistic features and grammatical features have on teaching and on direct and semi-direct speaking tasks.

### **Research Questions**

The study is intended to address the following questions:

1. To what extent do participants' performances on the direct and semi-direct speaking tests differ in terms of fluency, stress and tone choice?
2. To what extent do participants' performances on direct and semi-direct speaking tests differ in terms of grammatical analysis?
3. To what extent do participants' performances on direct and semi-direct speaking tests differ in terms of lexical analysis?

### **Methods**

#### **Procedures**

**Test administration and recording procedures.** The researcher briefly explained the study to each individual test taker, asked them to fill out a background questionnaire, and signed the consent form. They then entered a room where the interlocutor was waiting. The room was a classroom in the PIE building, and it was quite, as only the interlocutor and one test taker were inside the room at a given time. The test taker sat with the interlocutor, and the interlocutor conducted a 1-minute warm-up conversation with the participant. The interlocutor then made sure that the examinees understood the whole procedures of the speaking tests, and started the test. The interlocutor recorded the speech samples using an iPhone device. The test taker either started the semi-direct assessment first or the direct speaking assessment first. The interlocutor

was not allowed to talk to the participant once the recording had started.

**Interlocutor training.** Regarding the direct speaking assessment, the recruited interlocutors were trained to give instructions on the tests to the test takers. The interlocutors were asked to show the prompt card to the test takers and read the instructions (See Appendix) to the test takers. The instruction given by the interlocutor was as follows: “You will now be asked a question about a familiar topic. After you hear the question, you will have 15 seconds to prepare your response and 45 seconds to speak” (ETS, 2016). After the test takers completed the first section of the test, the interlocutor asked the test takers to perform another section using a computer, and the interlocutor recorded the speech samples using an iPhone device.

### **Linguistic Analysis**

All the speech samples were converted to WAV format, as the originally format created on the iPhone device was M4A. In order to perform the acoustical analysis, 24 speech samples were transcribed using Brazil’s (1997) model of intonation structure. The measure of intonation was then analyzed through PRAAT for the semi-direct and direct speaking assessments.

**Phonological measures.** The calculation measures of 10 acoustic variables are described in detail below.

Run/Tone Unit:

1. *Runs*: the utterances that pause 0.1 seconds and above. This analysis uses one run as a tone unit since the test takers had relatively low English proficiency.

Stress:

2. *Number of prominent syllables per run (Pace)*: it is calculated by counting the total number of prominent syllables and dividing them by the total number of runs.

## Pitch:

3. *High tone choice*: Based on the rising-pitch contour, high tone choice is identified on prominent syllables
4. *Level tone choice*: Based on the level-pitch contour, level tone choice is identified on prominent syllables
5. *Low tone choice*: Based on the decreasing pitch contour, low tone choice is identified on prominent syllables
6. *Overall pitch range*: This is calculated by the fundamental frequency of the highest prominent syllable minus the fundamental frequency of the lowest prominent syllable.

## Rate:

7. *Syllables per second*: Participants are supposed to produce two 45-second speech files; however, some of them spoke for less than 45 seconds. Therefore, this measure is calculated by the total number of syllables divided by the actual time it took to produce them.
8. *Mean length of runs*: This is calculated by the total number of syllables divided by the total number of runs.

## Pause:

9. *Number of silent pauses*: This is counted by the total number of pauses which that are 0.1 seconds or higher.
10. *Mean length of pauses*: This is calculated by the sum of pauses that are 0.1 seconds or higher and divided by the total number of pauses that are 0.1 seconds or higher.

**Grammatical measures.** The operationalization of grammatical features used is summarized below.

11. *Global accuracy*: This is calculated by counting the error-free T-units and dividing them by

the total number of T-units.

12. *Grammar complexity*: This is calculated by counting the verb phrases and dividing them by the total number of T-units.

**Lexical Measures.** The lexical feature is operationalized by token ratio.

13. *Token ratio*: The speech sample transcriptions are submitted to Vocabprofile website and performed through the classical option, which includes the General Service List (GSL) and Academic Word List (AWL).

### **Statistical Analysis**

Statistics included the mean, mode, median, skewness, and kurtosis. According to the descriptive statistics, some variables are normally distributed, while others are not normally distributed. For the normally distributed variables, paired t-tests were run; for the variables that are not normally distributed, Wilcoxon signed ranks tests were performed. The normality of variables are further discussed in the results session. The parametric/nonparametric statistical tests were run through SPSS separately. By comparing the means of the two sets of scores, the researcher could infer the variables that produced statistically significant variance.

### **Results**

The researcher examined the linguistics, grammatical, and lexical features of direct and semi-direct speaking assessments. The researcher conducted the grammatical analysis according to the grammar's accuracy and complexity, and calculated the acoustic measures by 10 variables. Finally, lexical measures were compared according to the VocabProfile website. Ten speech files were manually coded and are presented below; T-tests were then performed for each variable. (26 speech files were collected from 13 participants; the current analysis only covers five participants and 10 speech files).

### Fluency, Stress, and Pitch Measures

As shown in Table 1, 10 linguistic measures were described. According to the raw data, the runs exhibited in both delivery modes were close, given that the semi-direct speaking assessment presented more runs than the direct speaking assessment. The values in the stress measure show that participants produced more prominent syllables in the direct speaking assessment. The analysis included three tone choices. The percentages of rising tones were almost identical, but more level tones were observed in the semi-direct speaking assessment. In addition to this, the overall pitch range was larger in the direct speaking assessment than that in the semi-direct speaking assessment. The rate measures show that participants spoke faster in the face-to-face speaking assessment. The pause measures indicate that the computer-based assessment produced more silent pauses than the face-to-face one. However, by just looking at the raw data, one cannot infer that whether the comparison is statistically significant. Therefore, the researcher proceeded to perform parametric and nonparametric paired tests.

Table 1

#### *Acoustic Measures of Direct and Semi-Direct Speaking Assessments*

Acoustic/temporal measures		Face-to-face	Computer
Acoustic Measures			
1. Runs	Run/Tone unit	10.20	10.60
2. Prominent syllables per run (pace)	Stress Measure	1.31	0.97
3. Rising tone choice	Pitch Measures	0.35	0.35
4. Level tone choice		0.45	0.52
5. Fall tone choice		0.20	0.13
6. Overall pitch range		45.41	32.73
7. Syllables per second	Rate Measures	2.17	1.97
8. Mean length of Runs		10.11	7.96
9. Number of silent pauses/min	Pause Measures	9.20	9.60
10. Mean length of pauses		1.06	1.17



Descriptive statistics were then analyzed through SPSS, and the skewness and Kurtosis values of prominent syllables per run, the rising tone, number of pauses per minute, and mean length of pauses were not within the range of normal distribution. Hence, the aforementioned variables were then ran through the nonparametric Wilcoxon Signed Ranks Tests as seen in Table 2. As shown in Table 2, all of the  $p$  values are larger than the alpha level at 0.05, which indicates that these four variables under the direct and semi-direct speaking assessments' conditions do not make any difference.

Table 2

*Wilcoxon Signed Ranks Test for Acoustic Measures*

	prominent syllables per run	rising tone	number of pauses per minute	mean length of pauses
Sig. (2-tailed)	0.08	0.89	0.69	0.50

The rest of variables were calculated through paired t-tests, as seen in Table 3, because the same of group of participants took the semi-direct speaking assessment and direct speaking assessment.

Table 3

*Paired t-tests for Acoustic Measures*

	N	Sig.
Runs	5	0.13
Level Tone Choice	5	0.21
Fall Tone Choice	5	0.99
Overall Pitch Range	5	0.01
Syllables per Second	5	0.86
Mean Length of Runs	5	0.64

As presented in Table 3, after performing the t-tests, only the  $p$  value of the overall pitch

range,  $p=0.01$ , was smaller than the alpha level at 0.05, and it was statistically significant. Looking back at Table 1, one could infer that the direct speaking assessment results in a wider overall pitch range than the semi-direct speaking assessment.

Overall, based on the statistical tests, only the overall pitch range of the 10 acoustic measures exhibit a statistical difference across two testing delivery modes. The rest of the measures are close to each other.

### **Grammar Accuracy and Complexity**

To further examine the differences across two delivery modes, grammatical analysis were introduced, namely of the global accuracy and grammar complexity. The raw data are summarized in Table 4. Because the descriptive statistics of the grammar accuracy were not normally distributed, a Wilcoxon Signed Ranks Test was run for grammar accuracy, and a paired t-test was performed for grammar complexity.

Table 4

#### *Grammar Accuracy and Complexity*

	Face-to-face	Computer
Grammar Accuracy	6.00%	33.00%
Grammar Complexity	3.44	2.23

*Note.* Grammar accuracy was calculated according to the percentage of error-free T-units in the total number of T-units. Grammar complexity was calculated according to the verb-phrase ratios.

Grammar accuracy was rather low in the face-to-face speaking test at 6% compared to 33% in the computer-based test. However, the amount of verb phrases in the face-to-face test outnumbered the ones in the computer-based test.

Table 5

*Wilcoxon Signed Ranks Test for Grammar Accuracy*

Global accuracy	
Sig. (2-tailed)	0.08

As presented in Table 5, the  $p$  value of global accuracy, namely the ratio of error-free T-units of the total number of T-units, is  $p=0.08$ , which is larger than the alpha level at 0.05, which indicates that the grammar accuracy is not significantly different between the face-to-face and computer-mediated speaking assessments.

Table 6

*Paired t-test for Grammar Complexity (Verb Phrase Ratio)*

Verb Phrase Ratio	
Sig. (2-tailed)	0.834

As shown in Table 6, the verb phrase ratio, which is a potential indicator of grammar complexity, does not show any significant difference between the direct, and semi-direct speaking assessments because  $p = 0.83$ , which is larger than the alpha level at 0.05.

In summary, in terms of the grammatical analysis, there is not much difference between the direct and semi-direct speaking assessments.

**Lexical Analysis**

As shown in Table 7, the token ratio analysis was conducted in the Vocabprofile website. According to the raw data, participants in the face-to-face test produced more K-2 words than in

the computer-based test. On the other hand, participants produced more K-1 words in the computer-based test. Based on the descriptive statistics of both tests (See Appendix D), K-1 is not normally distributed, and K-2, AWL, and off-list words are normally distributed. Thus, parametric and non-parametric tests were performed, respectively.

Table 7

*Token Ratio in the VocabProfile*

Token (%)	F2f	Computer
K-1	89.83%	94.52%
K-2	2.61%	0.87%
AWL	2.72%	1.61%
Off-list	4.43%	3.00%

As shown in Table 8, the  $p$  value of K-1 is  $p=0.04$ , which is smaller than the alpha level at 0.05, which shows that the K-1 words produced in the direct speaking assessment and computer-mediated assessment are significant different. The raw data provides evidence that the computer-mediated speaking assessment's producing of more K-1 words does not occur by chance.

Table 8

*Wilcoxon Signed Ranks Test for K-1*

	K-1
Sig. (2-tailed)	0.04

However, as shown in Table 9, none of the three variables indicate a significant difference across the two testing modes because all of their  $p$  values are larger than the alpha level at 0.05.

Table 9

*Paired t-tests for K-2 and AWL*

	N	Sig.
K-2	5	0.462
AWL	5	0.806
Offlist	5	0.177

To summarize, in the lexical analysis, the only significant difference between the two delivery conditions is the numbers of K-1 words produced. Therefore, there is some evidence words.

### **Relevance to PIE and Second Language Learning**

The most prevalent way of assessing speaking in the Program of Intensive English (PIE) and/or in the country is still the face-to-face speaking assessment. Research shows that the two speaking tests almost elicit equivalent linguistic, grammatical and lexical features. However, the results show that most variables do not differ much except in overall pitch range and coverage of the first 1,000 words from the General Service List (GSL). The direct speaking assessment produces a wider pitch range, and the semi-direct speaking assessment produces more of the first 1,000 words from the GSL. A wider pitch range refers to a more animated voice as opposed to a monotonous voice. Therefore, the direct speaking assessment still has its advantages in terms of speaking assessment. Another implication is that participants tend to produce more common words in a semi-direct speaking assessment than in a direct speaking assessment.

### References

- Brazil, D. (1997). *The Communicative Value of Intonation in English Book*. Cambridge University Press.
- Cheng, W., & Warren, M. The use of rise and rise-fall tones in the Hong Kong Corpus of Spoken English.
- Eady, S. J. (1982). Differences in the F0 patterns of speech: tone language versus stress language. *Language and Speech*, Vol. 25.
- Educational Testing Service. (2014). Independent Speaking Rubrics. Retrieved from [https://www.ets.org/s/toefl/pdf/toefl\\_speaking\\_rubrics.pdf](https://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf)
- Educational Testing Service. (2016). TOEFL Practice Online. Retrieved from <https://toeflpractice.ets.org>
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, 39(3), 399-423.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, 18(03), 299-323.
- Hunt, K. W. (1970). Recent measures in syntactic development. In M. Lester (Ed.), *Readings in applied transformation grammar*. New York: Holt, Rinehart and Winston.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct?. *Applied linguistics*, 29(1), 24-49.
- Nakatsuhar, F., Inoue, C., Berry, V., Galaczi, E. (2016). Exploring performance across two delivery modes for the same L2 speaking test: Face-to-face and video-conferencing

- delivery- A preliminary comparison of test-taker and examiner behavior. Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/ielts-partnership-research-paper-1>
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64(4), 809-854.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*.
- Kang, O., & Pickering, L. (2013). Using acoustic and temporal analysis for assessing speaking. In A. Kunnan (Ed.), *Companion to Language Assessment* (pp.1047-1062). Wiley-Blackwell.
- Kenyon, D. M., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning & Technology*, 5(2), 60-83.
- Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.
- Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. Routledge.
- O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language testing*, 12(2), 217-237
- O'Loughlin, K. J. (2001). *The equivalence of direct and semi-direct speaking tests*. Cambridge University Press.
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom.

- TESOL Quarterly, 233-255.
- Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, 23, 19-43.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language testing*, 11(2), 99-123.
- Stansfield, C. W., & Kenyon, D. M. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. *System*, 20(3), 347-364.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(04), 905-916.
- Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.
- West, M., & West, M. P. (Eds.). (1953). *A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Addison-Wesley Longman Limited.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (No. 17). University of Hawaii Press.



Appendix

**Prompt Card**

Participant Code: \_\_\_\_001\_\_\_\_

Direct Speaking Assessment Task

Instruction: You will now be asked a question about a familiar topic. After you hear the question, you will have 15 seconds to prepare your response and 45 seconds to speak.

**Task: A friend of yours wants to go to university next year but cannot decide on a major field of study. What advice would you give your friend to help make this decision?**

Note:

Participant Code: \_\_\_\_002\_\_\_\_

Semi-Direct Speaking Assessment Task

Note: