Salient Linguistic Features of EFL Learner Spoken Corpus Elicited by a Computerized Speaking Test*

Young-Sun Moon (Korea University)
Inn-Chull Choi (Korea University)


This study attempted to identify the linguistic features (utterance, vocabulary, and syntactic) of English that are recognized as salient elements by native speakers. The oral responses of 60 EFL postsecondary learners comprised of six levels collected from a standardized computer-based speaking test were first transcribed to build a spoken corpus and were then analyzed for utterance, vocabulary, and syntactic features. In addition, scores of holistic and analytic scoring methods were compared. The findings were as follows: (1) Most of the indices of utterance, vocabulary, and syntactic structure demonstrated a significant difference between upper and lower proficiency groups. (2) The majority of indices had significant correlations with speaking proficiency, and the highest correlations with vocabulary indices. (3) The indices of all three categories accounted significantly for the variance of the speaking ability. (4) A very high correlation was found between holistic and analytic scoring methods. Based on the results that indicated which linguistic factors are considered significant by native speakers, meaningful implications could be drawn for effective methods of teaching speaking in the EFL environment focusing on vocabulary and willingness to communicate. Finally, the issue of comparability between holistic or analytic scoring methods was discussed.

**Key words**: computer-based oral proficiency interview, learner spoken corpus, linguistic features, corpus analysis, holistic scoring, analytic scoring

*doi: 10.15702/mall.2019.22.3.54*

I. INTRODUCTION

The theoretical framework for linguistic competence, which is the cornerstone of communicative competence, is well established in the work of Bachman (1990) and Bachman

* This research was supported by the College of Education, Korea University Grant in 2019.
and Palmer (1996). However, not much research has been conducted on how language competence interacts with language functions on the basis of empirical evidence, especially in the area of speaking assessment. This is presumably because of the fact that it is physically very difficult and time-consuming to construct a spoken corpus and analyze the quality of the spoken corpus after transcribing the collected speech data from a speaking test. In this context, investigating how language features, characteristics of utterance, and speaking language functions interact with each other may serve as an important research activity. In other words, analyzing multidimensional factors through the interaction of language elements and language functions will be a very meaningful research activity in terms of not only applied linguistics but also language acquisition theory.

If the salient features of speaking ability that are perceived as important to the native speaker can be identified, then that particular language ability can be considered an important element of communication ability and can help EFL students find an effective speaking learning strategy. Thus, the main objectives of the study are, first, to analyze the major indices of various utterance (speech rate, pausing, vocabulary/clause repetition, correction, etc.) and linguistic (vocabulary, and syntactic structure) characteristics. Next, by analyzing the relationship between linguistic traits, a more meaningful study can be conducted as a validity study of an English speaking performance assessment.

Another important issue in speaking assessment is related to scoring methods. Research has been conducted to determine which scoring method should be encouraged, but controversy over this topic still remains. Therefore, the purpose of this study is to analyze the spoken corpus of OPIc, which is one of the most widely used high-stakes computerized English speaking tests in Korea, to determine which factors of EFL speakers, in multidimensional terms such as speaking characteristics, vocabulary, and syntactic structure, are actually influential in the rating process of native speakers. This will help in forming an assessment framework for speaking tests. Additionally, holistic and analytic scoring methods are compared to choose scoring methods.

II. LITERATURE REVIEW

1. Speaking Ability

1) Componential Model of Communicative Language Ability
There are many factors that raters consider important when assessing speaking ability. However, it is often the case that the ability that significantly influences the speaking assessment process is not clear enough. In Orr's (2002) study, the University's Cambridge First Certificate in English Speaking test was used to report on scorers' thoughts when scoring grammar, vocabulary, discourse, pronunciation, and communication. The results of the study showed that the areas that the raters thought were important were not congruent with each other, and although the rating was based on specific rubrics, the subjective criterion for accepting the degree of error was also discrepant. This study suggests that the factors that are important in assessing speaking ability vary from rater to rater.

Further, there are some studies that have investigated the elements that play a pivotal role in the speaking assessment of Korean students (Chu, 2015; Kim, 2003). According to Kim (2003), there is no recognition of reliability or validity of assessment in many cases where teachers conduct English speaking assessment in secondary school. Since the majority of teachers lack theoretical background knowledge of speaking assessment, it is difficult to judge the factors that play a crucial role in assessing students' speaking ability in secondary schools. Chu (2015) analyzed 11 native speakers' assessment in an English speaking competition. The most significant factors in assessing students' overall speaking ability include fluency and content. According to the results of the evaluators' survey, fluency, content, and grammar were the most significant sub-factors in speaking as perceived by respondents. However, pronunciation did not appear to be an indicator of the assessment results, and Chu (2015) suggested that the unclear criteria for pronunciation was a reason for this result.

Although research on the analysis of the factors of speaking assessment is very important, the amount of literature on the topic is limited. This may be due to the fact that speaking skills are too elusive to be elaborately delved into in the EFL environment (Chuang, 2009), although they constitute a very integral part of language learning. Especially in Korea, there is a lack of education and assessment on speaking in the secondary education curriculum (Choi, 2008), and research in this area is even more lacking. This seems to be related to the fact that the CSAT has yet to incorporate speaking assessment. Additionally, transcribing is inevitable because of the nature of speaking research, and the fact that excessive time and expertise are required for transcribing is also one of the major reasons for the relatively little research on the components of speaking ability (O'Sullivan, Weir, & Saville, 2002).

In this respect, instead of identifying controversial issues of subskills or subcomponents of speaking proficiency, a componential model of language competence was proposed (Bachman, 1990; Bachman & Palmer, 1996). The communicative language ability consists of language...
competence, strategic competence, and psychophysiological skills (Bachman, 1990). Language competence is divided into organizational competence and pragmatic competence. Organizational competence is composed of grammar/accuracy and discourse competence, while pragmatic competence is composed of functional competence and sociolinguistic competence.

Although there are some differences in terminology, the four linguistic competences of this theoretical foundation are similar to the elements measured in the American Council on the Teaching of Foreign Languages (ACTFL; accuracy/comprehensibility, the text type (basic unit concept of discourse), the context/content, and communication functions (global tasks/functions)). It is worth noting that grammatical competence, or accuracy, includes pronunciation. In line with the theory of contemporary English education and assessment based on the framework of constructivism, this framework emphasizes the importance of overall comprehensibility over meticulous phonological phenomenon. Thus, five components, accuracy, discourse, functional competence, sociolinguistic (pragmatic) competence, and fluency (which is an operational manifestation of the strategic competence), constitute a comprehensive componential model for measuring speaking ability.

2) Features of Speaking Ability

Apart from the componential model defining the communicative language ability in general, several studies have attempted to identify the features of speaking ability by focusing on the speaking proficiency per se of language learners. The oral features of language learners can be principally categorized into linguistic (Iwashita, Brown, Menamara, & O’Hagan, 2008; Iwashita, 2010; Kang & Yan, 2018), cultural (Lee, 2007), and affective domains (Woodrow, 2006). In addition, the influence of features related to utterance such as pauses or hesitation on speaking performances have also been explored (Crowdy, 1995; Fletcher & Garman, 1995).

Among these features, linguistic components, including vocabulary, syntactic complexity, and grammatical accuracy, are known to greatly influence the speaker’s oral performance. This is because speakers must have the knowledge of the language in order to produce speech (Leong & Ahmadi, 2017). Thus, features that indicate the speaker’s use of vocabulary has been one of the prominent indicators of speaking proficiency. In addition, studies on speaking that examined syntactic complexity have mostly utilized the same measures frequently used in writing studies (Ortega, 1999). However, most of the studies have failed to systematically investigate the specific linguistic features of speaking performances that distinguishes language
learners’ speaking level.

One of the few major studies that examined whether features of speaking differed according to levels of proficiency is Iwashita et al. (2008), who found that the number of tokens, types, and unfilled pauses were the significant indicators (in grammar and vocabulary) that determined speaking levels. In addition, grammatical errors have been found to differentiate the proficiency levels based on speaking tasks, although adjacent levels did not show clear distinctions. However, the speech data of this study was based on monologic test tasks rather than real-life conversations or communication, thus diminishing the generalizability of the results of this study to target language use (TLU).

Another study done by Kang and Yan (2018) also investigated the features of speaking ability. In their study, 1 minute monologue speech files of 106 subjects were analyzed on the basis of quantitative and corpus analysis. The linguistic variables included in this study were related to fluency, coherence, grammatical accuracy, grammatical complexity, lexical resources, and pronunciation. Their study revealed that the difference of linguistic components was greater between high and low CEFR levels compared to adjacent levels. The pauses and hesitation markers decreased in higher levels and most of the lexical features in this study revealed obvious differences among the highest and lowest group. Nevertheless, this study also has its limitations in that it was based on relatively short utterances which may lead to misleading results.

2. EFL Speaker Language

English language learners have difficulties learning to speak in the EFL environment because English is mostly used for listening or reading. Ur (1996) asserted that among listening, reading, speaking, and writing, the most important ability is speaking because knowing a specific language means that one can speak it. Despite the importance of speaking ability, however, the very adverse nature of the EFL environment makes it difficult to acquire speaking ability effectively. The amount of input that learners receive in language acquisition is important, but without the output associated with communication, learning is not likely to lead naturally to the development of speaking or writing skills. This is because being able to comprehend the language is different from being able to produce it. It is generally agreed that speaking is the most difficult part for most EFL learners, and their English communication skills are still lacking (Zhang, 2009).

There have been studies on the factors that make it difficult for learners to learn speaking
(Dil, 2009; Littlewood, 1981; Ur, 1996). For instance, learners who are not yet proficient in English may have difficulty speaking due to lack of vocabulary, grammar, pronunciation, and oral skills (Al Hosni, 2014). The fact that speaking skills are not included in most English assessments in the EFL environment is also an important reason for speaking being neglected (Choi, 2008). Further, because the proportion of speaking in teaching methods and curricula is relatively small, students' speaking education may not be accomplished.

Ur (1996) listed four major motivation-related reasons for difficulty in improving speaking ability. Inhibition, low or uneven participation, and mother-tongue use can make it difficult for students to speak. Because of the fear and anxiety of being wrong, students may not be motivated to speak at all. Learner motivation is a very important factor, especially because students may not speak at all if they lack the motivation to (Littlewood, 1981). Students' overall participation rate may be low, and only students who actively speak may be involved. Additionally, because most students tend to use their mother tongue instead of using a foreign language, students' speaking within classrooms will be less frequent.

Considering that EFL learners are faced with an adverse linguistic environment due to various reasons, more effort is needed to maximize the overall efficacy of acquiring EFL by using more systematic methods of Teaching English as a Foreign Language (TEFL). Though it may not be feasible to deal with the elusive factors in affective and cognitive domains, it seems plausible to find an effective method to facilitate EFL learning by delving into salient linguistic features that exert impact on oral communication. In this respect, it would be desirable to identify significant features of good English communications skills perceived by competent native speakers and provide EFL learners with this information so that they can identify their problems with communicative skills by focusing on salient linguistic features and eventually improve their overall English proficiency. Therefore, the current study attempts to identify significant linguistic factors in speaking that indicate the level of speaking of Korean EFL learners across proficiencies based on various corpus indices while addressing the different scoring methods for speaking.

3. Scoring Methods

The scoring method of speaking test assessment can be roughly divided into two: holistic and analytic. The holistic method scores the candidate's utterance using a single overall score, while the analytic method uses different sub-scores to assess various aspects of the utterance (Iwashita & Grove, 2003). Overall, the holistic and analytic scoring methods both have
advantages and disadvantages as follows:

First, holistic scoring is the preferred method of speaking assessment to evaluate the effectiveness of overall communication (Weir, 1990). Holistic scoring is much more efficient and cost-effective than analytic scoring because the scoring process requires far less logistics including time and energy. In terms of the raters, it is also easier to score holistically. However, a disadvantage is its restriction to achieving the diagnostic purpose of the test with a single score. The more critical problem is that the relative importance of the sub-elements of speaking is unclear and arbitrary because it produces only one score (Xi, 2007).

Another possible problem is that if one considers a particular factor more salient according to subjective criteria, the validity of the score in holistic scoring will deteriorate. In other words, the score obtained by the holistic scoring method may not indicate the overall performance of speaking but may actually be the score of a specific part of speaking (Huot, 1990; Brown, Iwashita, McNamara, & O’Hagan, 2003). Furthermore, holistic scoring may be even more inappropriate for students who learn English as a second language (Hamp-Lyons, 1991). According to Carr (2000), in regard to writing assessment, learners who are nonnative speakers are more unfit to be scored in an integrated way because they show greater level differences across a wide range of linguistic domains than native speakers.

At the same time, unlike holistic scoring, the analytic scoring method has some merit in that it allows one to assess the relative importance of each factor by grading each score of the sub-elements (Xi, 2007). It is also desirable to use the analytic scoring method to include the different elements of communicative competence in the assessment process (Bachman & Savignon, 1986) because speaking is a complex function of various factors (Bachman & Palmer, 1982). In the aspect of scoring, interrater reliability tends to be higher because of the more detailed scoring of analytic methods (Hamp-Lyons, 1991). Additionally, analytic scoring is more useful in providing diagnostic evaluation information to learners (Carr, 2000; Choi, 2000). However, the analytic scoring method also has several drawbacks. First, it requires much more resources in terms of time and energy to score each of the sub-elements. The cognitive load of the raters is also an area of concern. Finally, it is very difficult to classify, define, and develop a clear rubric for scoring (Douglas & Smith, 1997).

Given the strengths and limitations of both scoring methods, the issue of which scoring method should be used for speaking assessment has been continuously discussed (Bachman, 1988; Bachman & Savignon, 1986; Chuang, 2009; Iwashita & Grove, 2003; Xi & Mollaun, 2006; Xi, 2007). Chuang (2009) found no statistically significant difference between holistic scoring and analytic scoring. However, Xi and Mollaun (2006) and Xi (2007), who analyzed the TOEFL
listening test, concluded that it is preferable to use the holistic scoring method rather than the analytic scoring method because of the small difference between the scores, which resulted in a high correlation between the three sub-factors in the analytic scoring method and an inability to give additional diagnostic information to learners. However, there have also been studies in favor of analytic scoring methods because they provide rich information about the language use of respondents, increase accuracy due to scoring based on clear criteria, and can adequately represent the multifaceted attributes of language skills (Brown & Bailey, 1984; Pollitt & Hutchinson, 1987). As there has been no conclusive study addressing the relationship between the two scoring methods and the analysis of linguistic qualities based on test takers' spoken corpus in an integrative manner, further research is required to delve into this controversial issue.

In pursuing these research objectives, the research questions were formulated as follows:

1. What is the extent to which utterance, vocabulary, and syntactic structure indices are related to and contribute to the speaking ability of EFL learners?
2. Based on the previous research question, what are the linguistic factors among utterance, vocabulary, and syntactic structure that native speakers perceive as significant?
3. What is the extent to which holistic scoring, analytic scoring, and significant linguistic factors are intercorrelated?

III. METHOD

1. Participants and Data

A total of 60 oral responses were extracted from a database of spoken corpus produced by a standardized English speaking test, OPIc (Oral Proficiency Interview Computer), administered in the manner of a computer-based oral proficiency interview in 2017 and 2018. As a simulated oral proficiency interview (SOPI) test, the OPIc is administered in such a way that test-takers are seated in a computer-based booth where they respond to a variety of oral questions and tasks presented by a virtual interviewer on the computer screen. Oral responses to 5 to 10 prompts (depending on different levels of proficiency) addressing topics of ordinary daily life (self-introduction, hobby, family, school major, occupation, etc.) were recorded and rated holistically by two experienced native raters per each test-taker. The rating was based on
the four elements, accuracy/comprehensibility, text type, context/content, and communication functions measured in the ACTFL scoring guidelines for speaking. Among the 9 different proficiency levels measured by OPIc, 6 levels (10 subjects per each level), ranging from intermediate to advanced were included in this study due to the availability of data. The levels are Intermediate Low (IL), Intermediate Mid1 (IM1), Intermediate Mid2 (IM2), Intermediate Mid3 (IM3), Intermediate High (IH), and Advanced Low (AL). Each subject's oral response lasted approximately 15 to 20 minutes in total.

The participants were divided into two groups according to their OPIc level in order to see whether the corpus features of their speaking were significantly different. The respondents were divided into two groups instead of three due to the limited sample size. Each level consisted of 30 respondents based on their OPIc level. The upper level was composed of IM3, IH, and AL levels, while the lower level was composed of IL, IM1, and IM2 levels.

To probe the comparability between holistic and analytic scoring, two experienced EFL experts (a bilingual rater and a native rater) were employed to measure the five components, accuracy (AC), discourse (DC), functional competence (FC), socio-pragmatic competence (SP), and fluency (FL).

2. Transcribing Process

Because the transcribing process is seldom reported on in research papers, the present study describes the process, which plays a critical part in scrutinizing the spoken corpus, in detail. Two EFL experts (trained by TEFL certification programs) participated in the transcribing process, which was supervised by a researcher/professor experienced in a series of EFL transcribing work and its subsequent research.

1) Transcribing the Oral Responses

The transcription process for building the spoken corpus of the responses for the computerized speaking test involved three steps. In the first step, all utterances, including both content and speech errors of the 10 responses at 6 levels, were transcribed. The transcript aimed to accurately record the exact utterances of the responses by transcribing all repeated words, stutters, incomprehensible pronunciations, and incomplete sentences.

Then, the second step of the transcribing process focused on ensuring that the transcription did not miss any utterances. Specifically, symbols and transcriptions of repetitions, corrections,
fillers, and incomprehensible sounds were checked again to determine whether they were precise. Further, each pause time was measured for its length to distinguish short and long pauses. Pause times were recorded up to five seconds because pauses longer than five seconds were not considered to be significantly different.

Finally, for the third and final transcription, the start and end times of the responses were recorded to calculate the response time, and the previous transcriptions were double-checked. The symbols used for the transcription are shown in Table 1, and examples of the transcription, including symbols, are also given.

<table>
<thead>
<tr>
<th>Type</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition</td>
<td>( )</td>
</tr>
<tr>
<td>Correction</td>
<td>( ( ) )</td>
</tr>
<tr>
<td>Filler</td>
<td>As pronounced (e.g., uh, um)</td>
</tr>
<tr>
<td>Stutter</td>
<td>[stutter]</td>
</tr>
<tr>
<td>Incomplete sentence</td>
<td>^</td>
</tr>
<tr>
<td>Incomprehensible pronunciation</td>
<td>[IC]</td>
</tr>
<tr>
<td>Pause time</td>
<td>[P1]</td>
</tr>
</tbody>
</table>

Example Sentences

1. (When I lived in there), when I lived in there, um, ((I had an)), ((I had a book shelf)), I had book shelf.
2. ...but after I moved into my current house, ((my fa[stutter])), (my parents), [P2] my parents...
3. ...I lived in a [IC] house with my family.
4. I just post some my^  

In the first example above, the respondent simply repeated the phrase 'when I lived there', so it was marked with ( ). Next, the word 'um', which is a filler, was recorded as well. The utterance before the final words, 'I had book shelf', was a correction and was marked as '((I had an)), ((I had an book shelf)).'

The second example shows the case of stuttering. The criterion for distinguishing stuttering and repetition/correction was whether the word was fully pronounced. The part where the respondent stuttered was marked as 'my fa[stutter].' Further, the pause time between 'my parents' was measured as two seconds and was recorded as [P2].

In the third example, the symbol for incomprehensible pronunciation [IC] was used because the utterance was not recognized as an audible word. In the fourth example, the respondent did
not complete the sentence during the utterance but ended it in the middle and passed on to the
next sentence.

2) Refining Process of the Transcription

After going through the three-step transcription process, the text files were refined for
further analysis. Text files including all the symbols were made, while separate files were created
without the symbols to analyze only the actual utterances of respondents.

First, all symbols including parentheses and the words used within symbols (parentheses (),
^, [IC], [P seconds], [stutter]) were deleted. Then, all fillers, repetitions, and corrections were
excluded and made into a separate file. The following is an example of the refining process (see
Table 2). The original files were intended to measure the frequency of utterance error types
such as repetition and correction. The text files from the first refining step were used for
vocabulary analysis and included as utterances of the respondent. Meanwhile, the text files from
the second refining step were utilized for syntactic structure analysis.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Refining Process of the Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
<td>Text</td>
</tr>
<tr>
<td>Original</td>
<td>Well, (if you), (when you), (when you enter the pub), (when you enter), when you enter the pub, the first thing you can see is a C shaped bar that ([I all]), I usually get all kinds of drinks.</td>
</tr>
<tr>
<td>First step</td>
<td>Well, if you, when you, when you enter the pub, when you enter, when you enter the pub, the first thing you can see is a C shaped bar that I all, I usually get all kinds of drinks.</td>
</tr>
<tr>
<td>Second step</td>
<td>Well, when you enter the pub, the first thing you can see is a C shaped bar that I usually get all kinds of drinks.</td>
</tr>
</tbody>
</table>

3. Utterance Indices

Indices for utterance were measured by calculating the frequency of various utterance
characteristics (Crowdy, 1995; Fletcher & Garman, 1995). First, the response time of each item
was calculated using the start and end times of the respondents and was then added to the
utterance time of all the questions. Additionally, the number of words per minute (WPM) was
used to represent the speaking speed and was calculated using both the original text and refined
text. The original WPM indicated the actual speed of speech, whereas the refined version of
WPM was the speed of speech without fillers. Next, the frequency of pause times, ranging from
one to five seconds, was also measured. All pauses longer than five seconds were recorded as
five seconds. Other common utterance errors such as filler, repetition, correction, stutter,
incomplete sentences, and incomprehensible pronunciation were also measured in this study.

4. Corpus Analysis

Corpus analysis was conducted to examine the vocabulary and syntactic structure of the respondents' utterances. Because vocabulary and grammar are very important factors in speaking, various indices were included.

1) Vocabulary Indices

For all the vocabulary indices except for the total number of words, which was based on the original transcription, analysis results from the lexical complexity analyzer (LCA) were used, which provided 33 indices (Ai & Lu, 2010; Lu, 2012). Because the LCA analysis site only supports the text file format, the transcript files that underwent the first refining process were saved in a text file format and analyzed. However, a handful of indices that were not found to have a significant relationship with the speaking scores or were relatively unimportant in the scope of this research based on previous studies were excluded. Indices such as total number of types, sophisticated word types, lexical types, and tokens were used to address length, while indices such as lexical density, lexical sophistication, and type-token-ratio (TTR) were included to indicate different lexical characteristics.

2) Syntactic Structure Indices

Next, for syntactic structure indices, analysis results of the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010) were used. Syntactic complexity indices consist of several types, including length according to utterance unit, complexity of sentence, and degree of dependency. L2SCA is also analyzed with text files, so the text files from the second refining process were used for analysis. Like the vocabulary indices, only syntactic structure indices that showed a significant relationship or were considered important based on the previous literature were included. For length, indices such as the total number of sentences, clauses, t-units, and complex nominals were included, and for syntactic complexity, mean length of sentence, mean number of clauses per sentence, etc., were used.
5. Statistical Analysis

As described above, various analysis tools were utilized to obtain the indices that indicated utterance, vocabulary, and syntactic structure characteristics of the speaking test responses. For statistical analysis, the respondents’ levels, IL, IM1, IM2, IM3, IH, and AL, were entered as numbers 1 through 6. After that, various analyses were conducted using data of all indices. The programs used for the analysis were Excel and SPSS 22. After using Excel to organize the data and sort them by category, SPSS was used. First, the descriptive statistics of the indices of each category were obtained by SPSS, and then t-test, correlation analyses and regression analyses were conducted with the indices of each category.

IV. RESULTS

1. Descriptive Statistics

1) Utterance Indices

The descriptive statistics of the utterance indices are shown in Table 3 below. The mean value of the response time was 19.20 (19 minutes and 12 seconds when converted into time), and the standard deviation was approximately 4 minutes. The WPM calculated with the total utterance was 97.54, which means the respondents uttered about 98 words on average in one minute. Meanwhile, the WPM calculated with the refined text files was 80.03, meaning that 80 meaningful words were spoken on average in one minute. The frequency of pause time tended to decrease as the pause time became longer, indicating that respondents frequently paused shortly. Among the frequency of the other indices, fillers were most frequent, followed by repetition, correction, stutter, incomplete sentence, and incomprehensible pronunciation. The kurtosis of certain indices indicated that the data did not follow a normal distribution, which may be due to the fact that utterance errors were found to be substantially more frequent in certain level respondents’ utterances.

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response time</td>
<td>19.20</td>
<td>3.99</td>
<td>10.15</td>
<td>26.05</td>
<td>-.85</td>
<td>.10</td>
</tr>
<tr>
<td>Words per minute (WPM;)</td>
<td>97.54</td>
<td>29.21</td>
<td>49.21</td>
<td>216.45</td>
<td>1.10</td>
<td>3.41</td>
</tr>
</tbody>
</table>

(Table 3) Descriptive Statistics for Utterance Indices
### Table 4: Descriptive Statistics for Vocabulary Indices

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of words</td>
<td>1921.18</td>
<td>821.14</td>
<td>585</td>
<td>5137</td>
<td>.98</td>
<td>2.64</td>
</tr>
<tr>
<td>Type</td>
<td>379.85</td>
<td>101.05</td>
<td>181</td>
<td>601</td>
<td>-.04</td>
<td>-.38</td>
</tr>
<tr>
<td>Sophisticated word type</td>
<td>99.33</td>
<td>34.46</td>
<td>36</td>
<td>188</td>
<td>.44</td>
<td>.22</td>
</tr>
<tr>
<td>Lexical type</td>
<td>308.13</td>
<td>91.33</td>
<td>132</td>
<td>512</td>
<td>.04</td>
<td>-.29</td>
</tr>
<tr>
<td>Sophisticated lexical type</td>
<td>93.52</td>
<td>33.75</td>
<td>33</td>
<td>181</td>
<td>.46</td>
<td>.30</td>
</tr>
<tr>
<td>Token</td>
<td>1661.50</td>
<td>722.28</td>
<td>475</td>
<td>4017</td>
<td>.66</td>
<td>.80</td>
</tr>
<tr>
<td>Sophisticated word token</td>
<td>187.10</td>
<td>77.97</td>
<td>50</td>
<td>446</td>
<td>.67</td>
<td>1.14</td>
</tr>
<tr>
<td>Lexical token</td>
<td>812.55</td>
<td>342.74</td>
<td>249</td>
<td>1911</td>
<td>.60</td>
<td>.75</td>
</tr>
<tr>
<td>Sophisticated lexical token</td>
<td>171.38</td>
<td>73.17</td>
<td>47</td>
<td>423</td>
<td>.74</td>
<td>1.42</td>
</tr>
<tr>
<td>Lexical density</td>
<td>.49</td>
<td>.02</td>
<td>.45</td>
<td>.53</td>
<td>-.12</td>
<td>-.54</td>
</tr>
</tbody>
</table>
Lexical sophistication .21 0.03 .16 .28 .45 -.24
Type-token-ratio (TTR) .25 0.06 .15 .41 .82 .26
Corrected TTR 6.71 0.66 5.56 8.43 .18 -.17
Lexical word variation .39 0.07 .20 .59 .38 .35
Verb variation 26.71 7.85 8.80 44.20 .07 -.42
Noun variation .44 0.07 .32 .64 .76 .50

3) Syntactic Structure Indices

The descriptive statistics of syntactic structure indices are shown in Table 5 below. The average number of sentences was 146.87, and the average length of sentences was 11.04 words. Further, the average number of t-units was 157.78, and the average length was 10.25 words. The average number of complex nominals, which is a common feature of logical writings and claims, was 146.17. Additionally, the syntactic structure of the respondents’ utterances was examined through the various indices given in the table.

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence</td>
<td>146.87</td>
<td>50.73</td>
<td>61</td>
<td>344</td>
<td>.97</td>
<td>2.83</td>
</tr>
<tr>
<td>Verb phrase</td>
<td>280.60</td>
<td>129.71</td>
<td>78</td>
<td>773</td>
<td>1.08</td>
<td>2.53</td>
</tr>
<tr>
<td>Clause</td>
<td>225.07</td>
<td>101.91</td>
<td>68</td>
<td>617</td>
<td>1.14</td>
<td>2.76</td>
</tr>
<tr>
<td>T-unit</td>
<td>157.78</td>
<td>54.77</td>
<td>63</td>
<td>363</td>
<td>.97</td>
<td>2.69</td>
</tr>
<tr>
<td>Dependent clause</td>
<td>69.25</td>
<td>46.62</td>
<td>9</td>
<td>210</td>
<td>.88</td>
<td>.39</td>
</tr>
<tr>
<td>Complex T-unit</td>
<td>50.87</td>
<td>30.28</td>
<td>8</td>
<td>138</td>
<td>.70</td>
<td>.14</td>
</tr>
<tr>
<td>Complex nominal</td>
<td>146.17</td>
<td>73.68</td>
<td>28</td>
<td>344</td>
<td>.48</td>
<td>-.34</td>
</tr>
<tr>
<td>Mean length sentence</td>
<td>11.04</td>
<td>2.26</td>
<td>6.77</td>
<td>18.06</td>
<td>.48</td>
<td>.57</td>
</tr>
<tr>
<td>Mean length T-unit</td>
<td>10.25</td>
<td>1.96</td>
<td>6.91</td>
<td>16.54</td>
<td>.65</td>
<td>.67</td>
</tr>
<tr>
<td>Mean length clause</td>
<td>7.40</td>
<td>.77</td>
<td>5.90</td>
<td>10.19</td>
<td>1.20</td>
<td>2.68</td>
</tr>
<tr>
<td>Clauses per sentence</td>
<td>1.49</td>
<td>.26</td>
<td>1.03</td>
<td>2.15</td>
<td>.23</td>
<td>-.37</td>
</tr>
<tr>
<td>Clauses per T-unit</td>
<td>1.39</td>
<td>.22</td>
<td>.97</td>
<td>2.02</td>
<td>.52</td>
<td>-.06</td>
</tr>
<tr>
<td>Dependent clause ratio</td>
<td>.28</td>
<td>.09</td>
<td>.09</td>
<td>.48</td>
<td>.08</td>
<td>-.54</td>
</tr>
<tr>
<td>T-units per sentence</td>
<td>1.08</td>
<td>.05</td>
<td>.95</td>
<td>1.21</td>
<td>-.13</td>
<td>.06</td>
</tr>
<tr>
<td>Complex T-unit ratio</td>
<td>.30</td>
<td>.12</td>
<td>.08</td>
<td>.60</td>
<td>.27</td>
<td>-.27</td>
</tr>
<tr>
<td>Complex nominals per T-unit</td>
<td>.89</td>
<td>.27</td>
<td>.39</td>
<td>1.77</td>
<td>.77</td>
<td>1.00</td>
</tr>
<tr>
<td>Complex nominals per clause</td>
<td>.63</td>
<td>.12</td>
<td>.35</td>
<td>1.09</td>
<td>.74</td>
<td>2.43</td>
</tr>
</tbody>
</table>
2. Independent Samples *t*-test

To determine whether there were differences in linguistic features between the upper (IM3, IH, AL) and lower (IL, IM1, IM2) level respondents, an independent samples *t*-test was run.

First, the results of the *t*-test for the utterance indices are given in Table 6. The majority of indices showed a statistically significant difference except for repetition, incomplete sentence, and incomprehensible pronunciations. The significant difference of response time indicated that the upper level respondents’ utterance was significantly longer (*t* = 4.939, *p* < .01). The WPM of both total (*t* = 4.720, *p* < .01) and refined (*t* = 4.471, *p* < .01) texts also had a significant difference. Pause times longer than one second were more frequent in the lower group, while shorter pause times were more frequent in upper level respondents. Also, the number of fillers (*t* = 2.518, *p* < .05), corrections (*t* = 4.426, *p* < .01), and stutters (*t* = 5.249, *p* < .01) were more frequent in the upper level utterances, showing that they utilized these utterance strategies more often.

![Table 6](image)

Next, the *t*-test results for the vocabulary indices are given in Table 7. All of the indices except for lexical sophistication and corrected TTR demonstrated a significant difference. In
particular, the length-related indices, such as the total number of words \((t = 5.994, p < .01)\), type \((t = 5.422, p < .01)\), and token \((t = 5.802, p < .01)\) were all significantly different between the upper and lower group. It is notable that the lower group had a significantly higher lexical density \((t = -2.687, p < .01)\), TTR \((t = -5.093, p < .01)\), lexical word variation \((t = -3.751, p < .01)\) and noun variation \((t = -4.284, p < .01)\) than the upper level group. However, these results do not necessarily indicate that the lower level respondent’s utterances had a lexically more diverse response due to the fact that these indices are sensitive to utterance length.

\[
\begin{array}{llllll}
\text{Index} & \text{Upper 50%} & \text{Lower 50%} & \text{t-test} \\
\hline
\text{Total number of words} & 2424.77 & 755.20 & 1417.60 & 526.12 & 5.994** \\
\text{Type} & 437.97 & 84.12 & 321.73 & 81.92 & 5.422** \\
\text{Sophisticated word type} & 119.37 & 30.49 & 79.30 & 25.62 & 5.511** \\
\text{Lexical type} & 360.10 & 77.32 & 256.17 & 73.53 & 5.335** \\
\text{Sophisticated lexical type} & 112.87 & 30.03 & 74.17 & 25.32 & 5.397** \\
\text{Token} & 2095.57 & 660.13 & 1227.43 & 485.58 & 5.802** \\
\text{Sophisticated word token} & 233.03 & 70.88 & 141.17 & 54.58 & 5.625** \\
\text{Lexical token} & 1013.60 & 312.87 & 611.50 & 239.74 & 5.587** \\
\text{Sophisticated lexical token} & 213.90 & 67.02 & 128.87 & 51.58 & 5.507** \\
\text{Lexical density} & .48 & .02 & .50 & .02 & -2.687** \\
\text{Lexical sophistication} & .21 & .03 & .21 & .03 & .413 \\
\text{Type-token-ratio (TTR)} & .22 & .04 & .28 & .06 & -5.093** \\
\text{Corrected TTR} & 6.83 & .56 & 6.58 & .74 & 1.427 \\
\text{Lexical word variation} & .35 & .06 & .42 & .07 & -3.751** \\
\text{Verb variation} & 28.88 & 6.28 & 24.54 & 8.73 & 2.208* \\
\text{Noun variation} & .40 & .04 & .47 & .08 & -4.284** \\
\end{array}
\]

Finally, the t-test results of the syntactic structure indices are given in Table 8 below. Similar to the other categories, most of the indices, except for the mean length of clause and the number of t-units per sentence, had a significant difference. All of the indices indicating the length of utterance on the syntactic level were significantly different between the two groups. Also, several indices related to t-unit such as the number of t-units \((t = 4.798, p < .01)\), complex t-unit \((t = 4.334, p < .01)\), mean length of t-unit \((t = 3.415, p < .01)\), and the complex t-unit ratio \((t = 3.044, p < .01)\) showed a significant difference between the upper and lower group.
The results of the t-test revealed that most of the indices were significantly different between the upper and lower group. This indicates that utterance, vocabulary, and syntactic indices may serve as indicators that distinguish the level of respondents.

3. Correlation Analysis

Next, correlation analyses were conducted for the participant’s level with each category, utterance, vocabulary, and syntactic structure. The results of the correlation analysis of the utterance indices showed that all utterance categories had a significant correlation with the respondents’ speaking ability level except for repetition, incomplete sentences, and incomprehensible pronunciations. The 10 indices that showed the highest significant correlations with speaking ability level were as given in Table 9. The response time of the utterance had the highest correlation \((r = .697, p < 0.01)\), followed by the different types of pause times and WPM. Repetition and stutter were also significantly correlated to respondents’
speaking ability level.

Second, all vocabulary indices except lexical sophistication showed a significant correlation, and the 10 indices that demonstrated the highest correlation are also given in Table 9. The majority of the indices were related to the length of the utterance. For instance, the number of tokens \( r = .793, p < .01 \) indicating the length of the refined response and the total number of words \( r = .784, p < .01 \) indicating the length of the original utterance were the indices with the highest correlations. Other indices except verb variations were also related to the number of different types of tokens and types, which in turn were related to the quantity (type) or quality (indices related to sophisticated words or lexical words) of the vocabulary used in speaking.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>( r )</th>
<th>Vocabulary</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Response time</td>
<td>.697</td>
<td>Token</td>
<td>.793</td>
</tr>
<tr>
<td>2 Pause time [P3]</td>
<td>.687</td>
<td>Total number of words</td>
<td>.784</td>
</tr>
<tr>
<td>3 WPM (Refined)</td>
<td>.678</td>
<td>Lexical token</td>
<td>.773</td>
</tr>
<tr>
<td>4 WPM (Total)</td>
<td>.662</td>
<td>Type</td>
<td>.772</td>
</tr>
<tr>
<td>5 Pause time [P1]</td>
<td>.593</td>
<td>Lexical type</td>
<td>.769</td>
</tr>
<tr>
<td>6 Pause time [P4]</td>
<td>.577</td>
<td>Sophisticated word type</td>
<td>.755</td>
</tr>
<tr>
<td>7 Pause time [P2]</td>
<td>.576</td>
<td>Sophisticated lexical type</td>
<td>.752</td>
</tr>
<tr>
<td>8 Pause time [P5]</td>
<td>.552</td>
<td>Sophisticated word token</td>
<td>.738</td>
</tr>
<tr>
<td>9 Repetition</td>
<td>.490</td>
<td>Sophisticated lexical token</td>
<td>.736</td>
</tr>
<tr>
<td>10 Stutter</td>
<td>.445</td>
<td>Verb variation</td>
<td>.411</td>
</tr>
</tbody>
</table>

** All correlations were significant at the .01 level (2-tailed)

Next, the majority of the syntactic structure indices showed a significant correlation, except for the mean length of clause and the average number of t-units in one sentence. The 10 indices that showed the highest correlation among the syntactic structure indices are given in Table 10 below. The complex nominal showed the highest correlation \( r = .796, p < .01 \), whereas several indices related to the t-unit, such as the number of complex t-units \( r = .730, p < .01 \) and the mean length of the t-unit \( r = .688, p < .01 \), were highly correlated with the respondent’s speaking ability level.

Finally, the 10 indices with the highest correlation coefficients regardless of category were sorted. Table 10 shows the ten most highly correlated indices across all categories. Except for the number of complex nominals in the syntactic structure category which demonstrated the
highest correlation, all nine indices were related to vocabulary, while no utterance indices were included. Nonetheless, the characteristics of the respondents’ utterances cannot be considered unimportant based on this result because the utterance indices also showed a relatively high correlation.

<table>
<thead>
<tr>
<th>TABLE 10</th>
<th>Syntactic Structure with Highest Correlation Coefficients and Correlated Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic Structure</td>
<td>r</td>
</tr>
<tr>
<td>1 Complex nominal (CN)</td>
<td>.796</td>
</tr>
<tr>
<td>2 Verb phrase</td>
<td>.736</td>
</tr>
<tr>
<td>3 Dependent clause</td>
<td>.731</td>
</tr>
<tr>
<td>4 Complex T-unit</td>
<td>.730</td>
</tr>
<tr>
<td>5 Clause</td>
<td>.727</td>
</tr>
<tr>
<td>6 Mean length T-unit</td>
<td>.688</td>
</tr>
<tr>
<td>7 T-unit</td>
<td>.668</td>
</tr>
<tr>
<td>8 Mean length sentence</td>
<td>.666</td>
</tr>
<tr>
<td>9 Clauses per T-unit</td>
<td>.663</td>
</tr>
<tr>
<td>10 CN per T-unit</td>
<td>.663</td>
</tr>
</tbody>
</table>

** All correlations were significant at the .01 level (2-tailed)  
** S Syntactic Structure, V Vocabulary

Most of the indices across all three categories were found to have a significant relationship with the speaking ability level of respondents, indicating that the indices may have been linguistic factors of speaking that the raters had in mind. In particular, the vocabulary indices were among the highest correlated indices.

4. Regression Analysis

1) Multiple Regression Analysis

After conducting the correlation analyses, regression analyses were conducted to determine which indices could significantly contribute to the speaking ability of respondents. First, multiple regression analysis was conducted with the 10 highest indices of each category given in the previous correlation analyses results tables.
As shown in Table 11 above, regression analyses of all categories showed significant results. The results of the utterance indices showed the highest explanatory power (\( R^2 = .827, F = 23.433, p < .001 \)). At the same time, vocabulary, syntactic structure, and indices with the highest correlation coefficients also showed significant predictability of 72.2%, 68.7%, and 71.7%, respectively (\( F = 14.451, p < .001; F = 10.778, p < .001; F = 12.433, p < .001 \)).

2) Multiple Hierarchical Regression Analysis with Utterance Indices

Next, multiple hierarchical regression analyses were conducted for each category to determine which indices accounted significantly for the speaking ability of the respondent. To begin with, the results of the multiple hierarchical regression analysis of the utterance indices are shown in the following Table 12. Four indices that showed a significant predictive power were entered in each step. The response time significantly accounted for 48.6% of the respondents' level when entered in the first step (\( \Delta F = 54.79, p < .001 \)). Then, when the pause time of three seconds was entered in the second step, it added 25.8% of the variance above and beyond the response time (\( \Delta F = 57.20, p < .001 \)). The WPM calculated with the refined text and pause time of one second showed an additional significant explanatory power when entered in the third and fourth steps of the regression analysis after the previous variables were controlled for (\( \Delta R^2 = .028, \Delta F = 6.78, p < .05; \Delta R^2 = .019, \Delta F = 4.93, p < .05 \), respectively). The multiple hierarchical regression analysis results showed that the four indices of utterance significantly accounted for 79% of the respondents’ levels. This indicated that utterance indices have a high degree of accountability for respondents’ speaking ability.
3) Multiple Hierarchical Regression Analysis with Vocabulary Indices

Next, the results of the multiple hierarchical regression analysis of the vocabulary indices are shown in Table 13. Three indices were entered in the regression model. The number of tokens was entered in the first step, and it explained 62.8% of the variance of the speaking ability level ($\Delta F = 97.967, p < .001$). Then, when the number of lexical tokens was entered in the second step, it explained an additionally significant predictive power when the number of tokens was controlled for ($\Delta R^2 = .036, \Delta F = 6.166, p < .05$). Also, the number of sophisticated word types added an additional 3.3% of variance above and beyond the previously entered variables in the third step of the regression analysis ($\Delta F = 6.073, p < .05$). The number of tokens demonstrated a strong predictive power for speaking ability level, whereas the number of lexical and sophisticated words showed significant explanatory power when entered after token. Together, the three indices related to vocabulary showed a total explanatory power of 69.7% for the level of respondents.

4) Multiple Hierarchical Regression Analysis with Syntactic Structure Indices

The results of the multiple hierarchical regression analysis of the syntactic structure indices are shown in Table 14. When regression analysis was conducted with the 10 indices of the syntactic structure category, only the number of complex nominals was a significant predictor for speaking ability level. Complex nominal alone significantly contributed to 63.4% the respondents’ speaking ability level ($\Delta F = 100.371, p < .001$), while the other indices failed to
demonstrate a significant predictive power.

\[ \text{(TABLE 14) Multiple Hierarchical Regression Analysis with Syntactic Structure Indices} \]

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>R</th>
<th>( R^2 )</th>
<th>( \Delta R^2 )</th>
<th>( \Delta F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Complex nominal</td>
<td>.796</td>
<td>.634</td>
<td>.634</td>
<td>100.371</td>
<td>.000</td>
</tr>
</tbody>
</table>

The results of the multiple hierarchical regression analysis with the 10 most highly correlated indices regardless of category showed that only the number of complex nominals was entered, which is the same result as the regression analysis of the syntactic structure indices. Therefore, the result was not included. This result indicated that complex nominals, which are a common syntactic structure feature in logical writings, had a powerful explanatory power.

To sum up, the results of the multiple regression analysis overall showed that the utterance indices had the greatest predictive power when the 10 indices were entered together, followed by vocabulary and syntactic structure indices. At the same time, the results of the hierarchical regression analysis showed specific indices that had significant predictive powers. For the utterance indices, the response time, pause times, and WPM of the refined text were significant predictors, while token, lexical token, and sophisticated word type were significant predictors among the vocabulary indices. Further, the number of complex nominals was the only significant predictor among both the syntactic structure indices and the 10 highest correlated indices among all categories.

5. Holistic Scoring vs. Analytic Scoring

Based on the findings of the correlation and regression analyses, it was considered well worth the effort to explore the extent to which the holistic scoring results correlated with the analytic scoring results. As explained in the third chapter, analytic scoring was conducted to measure the five components AC, DC, FC, SP, and FL.

Analytic scoring was conducted by two experienced EFL teachers (a bilingual rater and a native English rater). The interrater correlation coefficients were calculated for each component as follows: AC: .957; DC: .958; SP: .949; FC: .949; and FL: .958. Because the interrater correlations turned out to be very high, it was considered reasonable to obtain an average for each component and compare it with the holistic scoring results in further analyses.

It would be of great interest to English language testers and teachers to probe the interrelationship among the two scoring results and the aforementioned linguistic factors. The
results presented in Table 15 revealed very high correlations (all of which were over .9) between the holistic and analytic scoring results. This strongly suggests that holistic scoring may be a viable approach to measuring EFL speaking proficiency, if necessitated by restricted logistic conditions of performance testing. It is well documented that both scoring methods have their strengths and limitations in that the holistic scoring method definitely enhances practicality, while the more time-consuming labor-intensive analytic scoring method can provide test takers with more diagnostic information. This finding allows teachers and testers to justify the choice of an appropriate scoring method, given their varying testing conditions and the goals of teaching and testing.

The results also indicated that the 10 indices had fairly high correlations (all of which were around .8) with both holistic and analytic scoring, and that the analytic scores resulted in slightly higher correlations with all the indices (except for TW) compared to holistic scoring. Other noteworthy findings regarding the relationship between the scoring method were as follows: the complex nominal, which constituted the only grammatical index in the final stage of analysis, had higher correlations with both holistic and analytic scoring results than all the other indices. This implies that this grammatical concept merits slightly more attention from raters than other vocabulary-related aspects. It is also worth noting that the four sophisticated word-based indices had correlations with both scoring results at a level of less than .8. This may be due to the fact that the overall difficulty of sophisticated vocabulary produced in the speaking test contributed somewhat less to higher ratings than the other features.
TABLE 15  Correlation Analysis among Holistic and Analytic Scoring Components

<table>
<thead>
<tr>
<th>Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Tk</td>
<td>.971</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. TW</td>
<td>.937</td>
<td>.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. LTk</td>
<td>.968</td>
<td>.996</td>
<td>.966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. T</td>
<td>.930</td>
<td>.923</td>
<td>.880</td>
<td>.926</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. LT</td>
<td>.932</td>
<td>.925</td>
<td>.883</td>
<td>.930</td>
<td>.998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. SWT</td>
<td>.896</td>
<td>.888</td>
<td>.870</td>
<td>.897</td>
<td>.956</td>
<td>.959</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. SLT</td>
<td>.896</td>
<td>.887</td>
<td>.868</td>
<td>.896</td>
<td>.956</td>
<td>.960</td>
<td>.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. SWTk</td>
<td>.914</td>
<td>.941</td>
<td>.931</td>
<td>.954</td>
<td>.905</td>
<td>.909</td>
<td>.939</td>
<td>.937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. SLT</td>
<td>.917</td>
<td>.941</td>
<td>.931</td>
<td>.952</td>
<td>.907</td>
<td>.910</td>
<td>.943</td>
<td>.942</td>
<td>.995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. HScr</td>
<td>.796</td>
<td>.793</td>
<td>.784</td>
<td>.773</td>
<td>.772</td>
<td>.769</td>
<td>.755</td>
<td>.752</td>
<td>.738</td>
<td>.736</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. AAC</td>
<td>.825</td>
<td>.806</td>
<td>.781</td>
<td>.792</td>
<td>.813</td>
<td>.810</td>
<td>.791</td>
<td>.788</td>
<td>.764</td>
<td>.765</td>
<td>.966</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. ADC</td>
<td>.835</td>
<td>.811</td>
<td>.782</td>
<td>.798</td>
<td>.813</td>
<td>.810</td>
<td>.788</td>
<td>.784</td>
<td>.761</td>
<td>.762</td>
<td>.970</td>
<td>.991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. ASP</td>
<td>.830</td>
<td>.809</td>
<td>.783</td>
<td>.794</td>
<td>.809</td>
<td>.806</td>
<td>.785</td>
<td>.782</td>
<td>.757</td>
<td>.758</td>
<td>.976</td>
<td>.990</td>
<td>.996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. AFC</td>
<td>.831</td>
<td>.809</td>
<td>.783</td>
<td>.794</td>
<td>.807</td>
<td>.804</td>
<td>.781</td>
<td>.778</td>
<td>.755</td>
<td>.755</td>
<td>.973</td>
<td>.989</td>
<td>.996</td>
<td>.999</td>
<td></td>
</tr>
</tbody>
</table>

** All correlations were significant at the 0.01 level (2-tailed).
** CN: Complex Nominal; Tk: Token; TW: Total Word; LTk: Lexical Token; T: Type; LT: Lexical Type; SWT: Sophisticated Word Type; SLT: Sophisticated Lexical Type; SWTk: Sophisticated Word Token; SLTk: Sophisticated Lexical Token; HScr: Holistic Score; AAC: Analytic Score-Accuracy; ADC: Analytic Discourse; ASP: Analytic SocioPragmatic; AFC: Analytic Functional Competence; AFL: Analytic Fluency

V. CONCLUSION & IMPLICATIONS

The current study analyzed a spoken corpus extracted from OPIc to examine the linguistic characteristics of EFL test takers' utterances, vocabulary, and syntactic structure and their interrelationship with holistic and analytic scoring results. First, the overall characteristics of the speaking test responses were explored through the various utterance, vocabulary, and syntactic structure indices. Next, a significant difference for the indices was found between upper and lower level groups for the majority of indices. The correlation analyses indicated that most of the indices of all three categories had significant correlations with the speaking ability level. Meanwhile, the vocabulary indices demonstrated a relatively stronger relationship in comparison with utterance features and syntactic structure indices.
An additional analysis of multiple regression was conducted with the 10 highest correlated indices of each category. The utterance indices significantly accounted for the largest amount of variance of the speaking ability levels, followed by vocabulary and syntactic structure indices. Finally, multiple hierarchical regression analyses were conducted to identify the exact indices that had significant predictability. The overall indices that native speakers perceive as significant in speaking ability in general included response time, WPM, pause time, utterance length, vocabulary proficiency, and complex nominals, which were the significant indicators of the participants’ speaking level. The results were congruent with previous research that also found that utterance, vocabulary, and syntactic features of speaking were distinctive among different speaking levels (Iwashita et al., 2008; Kang & Yan, 2018).

In order to explore the relationship between the two different scoring methods, the final correlation analysis was conducted with holistic scoring, the 5 components of the analytic scoring method, and the 10 salient linguistic factors. The holistic score and all components of the analytic scoring method showed high correlations. Additionally, the linguistic factors demonstrated a high correlation with both scoring methods, whereas the components of analytic scoring had a higher correlation for the majority of indices.

The results of the current study revealed that the majority of indices of all categories were significantly different between upper and lower level groups, and were also significantly related to speaking ability, indicating that these indices may be the linguistic factors that influence the scoring process of native raters. Furthermore, the vocabulary indices demonstrated the highest correlation coefficients, suggesting that the total amount of utterances and vocabulary proficiency of the speaker has a stronger relationship with speaking ability than syntactic-structure-related indices. The regression analyses also showed that the utterance indices contributed the largest amount of variance, indicating that these linguistic factors are also important in the process of rating speaking.

Additionally, the fact that length-related indices showed high correlations suggested that being able to produce longer utterances and productive vocabulary (not that sophisticated) is perceived by educated native speakers as salient features of good speaking skills. This is in line with well-established psycholinguistic findings that semantic mapping plays a much more significant role than syntactic analysis in natural cognitive processing of linguistic stimuli. These results provide important implications for teaching speaking in EFL context; more focus should be placed on ordinary vocabulary and willingness to communicate (WTC).

As for the comparability of the two scoring methods, the high intercorrelations among holistic scoring, analytic scoring, and linguistic features suggested that both scoring methods
were appropriate for testers and teachers to choose from. Thus, the choice of scoring method should be based on consideration of the setting and the goal of assessment, including its inherent advantages and disadvantages.

This study is not without limitations, however, in that the insufficient sample size precludes the generalizability of the research findings. With more participants for each level of English speaking proficiency, more sophisticated factor-related analyses could have been conducted to obtain more in-depth explanations for underlying factors of speaking ability. Additionally, other linguistic factors such as phonological units (segmentals and suprasegmentals) were not addressed in this study. Also, one of the most important factors in speaking is the appropriateness of productive vocabulary usage. However, the degree of appropriateness was not measurable with the current analysis tools and were therefore excluded from the scope of this study. Including a wider range of variables representing various features of EFL learner's speaking proficiency may make it possible to explain the unexplained variance of the speaking ability.

Though ordinary native interlocutors are not able to analyze all linguistic features of speaking ability on a conscious level, well-trained raters are capable of perceiving significant features as indicators of EFL learners’ oral proficiency. Therefore, it would be well worth our effort to enhance the efficacy of TEFL by identifying significant linguistic features perceived (by competent native raters) as factors playing a pivotal role in determining the overall comprehensibility and proficiency. It is hoped that a further study will employ more data and cover more features of speaking, and more accurately identify the linguistic factors that are perceived as salient features and eventually provide helpful guidance for both valid assessment and effective acquisition of English communication skills.

REFERENCES


Applicable levels: secondary, tertiary education
Authors: Moon, Young-Sun (Korea University, 1st author); youngsun1993@korea.ac.kr
        Choi, Inn-Chull (Korea University, Corresponding author); icchoi@korea.ac.kr

Received: July 31, 2019
Reviewed: August 20, 2019
Accepted: September 15, 2019