



A Comparative Study of Using Analytic and Holistic Methods of Scoring in Measuring Speaking Skill Among Iranian Second-Year University EFL Students

Behrouz Ahmadi^a, Ehsan Namaziandost^{b*}, Parisa Etemadfar^c

^aDepartment of English, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

^bDepartment of English, Faculty of Humanities, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran

^cDepartment of English Language, Faculty of Letters & Humanities, Shahrekord University, Shahrekord, Iran

Article Info

Article history

Submitted: 30 September 2019

Accepted: 28 January 2020

Published: 2 March 2020

Keywords:

Oral skill evaluation; analytic scoring; holistic scoring; teaching English as a foreign language.

*Correspondence Address:

e.namazi75@yahoo.com

Abstract

Using analytic and holistic methods of scoring in measuring speaking skill was not investigated in Iranian EFL context. Therefore, in this study, the speaking skill of English belonging 70 individuals, those were Iranian second-year university EFL students, were evaluated by speaking addressee and examiner. The speaking addressee performs the holistic scoring though the examiner directed the analytic scoring. Context and construction, vocabulary, grammar and pronunciation are the classes of analytic scoring. The analytic mean of four scales was 3.396, although the mean of holistic scoring was 3.628. The results showed a statistically considerable diversity between analytic and holistic methods of evaluation considering that p-value was estimated at 0.002 ($P < 0.05$). Therefore, it is recommended that employing these two scoring techniques in the procedure of evaluation may be considered proper seeing that one augment the other and lead to more inclusive evaluation.

© 2020 English Education Department, University of Sultan Ageng Tirtayasa

INTRODUCTION

The speaking skill is critical among the four skills (Khamkhien, 2010) while people have adequate information relating to a language is mainly referred to as speakers of the particular language (Abedi, Keshmirshakan, & Namaziandost, 2019; Ur, 2012). Similarly, Pokrivčáková (2010) confirms that abundant foreign language trainers and students suppose speaking skills like the comprehending amount of a language. Göktürk (2016, p.71) furthermore attaches considerable

importance to speaking presentation: "with the proliferating significance belonging to speaking as part of one's language ability within the Communicative Language Teaching sample, the instructing of speaking skills in second language learning has become an enthusiastic zone of research over the past two decades". It is in like manner universalization and computerized time that performs an inconceivable instrument as influential oral communication skills have determined to be exceedingly essential in this time (Hashemifardnia, Namaziandost, & Rahimi Esfahani, 2018b; Murugaiah, 2016). Although concurrently, speaking might be regarded as the supreme disturbing skill to require as language has to be generated promptly and random, that requires a significant deal of exercise (Anderson, 2015; Nasri, Namaziandost, & Akbari, 2019). Undoubtedly, it takes long time and constant attempt for a foreign language learner to become skillful in the speaking skills.

For most EFL learners, learning English means being able to speak fluently. Language is basically oral. Therefore, speaking occupies an important place in any matter of teaching and learning foreign languages. For most people, the ability to speak a language is synonymous with knowing that language since speech is the most basic means of human communication" (Hashemifardnia, Namaziandost, & Rahimi Esfahani, 2018b, p. 8). Moreover, the ability to function in another language is generally characterized in terms of being able to speak that language. When someone asks, "Do you know another language?", they generally mean "Can you speak the language?". To teach speaking skill some old methods are used in our country which cannot develop the learners' ability to speak the target language fluently. The present study focused on analytic and holistic methods of scoring in measuring speaking skill among Iranian second-year university EFL students.

Concerning the evaluation of speaking skill, O'Sullivan (2012, p.234) affirms that "it is customarily believed that the most troublesome tests to expand and execute are tests of spoken language ability". In the same case, Chuang (2009) asserts as there are abundant inner and outer factors that influences on examiners, evaluating speaking presentation seems to be one of the utmost rigid duties to executive. As well, Luoma (2004) asserts that speaking evaluation is provoking due to the fact that there are abundant elements that influences on the conception of an examiner concerning how great an individual is able of speaking. In addition, examiners consider examine scores to be exact and proper

for the aims of estimating spoken skill, which is not eternally. Hence, performing proper and authentic evaluation of speaking presentation is a slightly rigid duty and requires a lot of features to be thought.

Approaches of Evaluating Speaking Skills

Two approaches named analytic and holistic scoring are applied to assess the oral skills which are ordinarily utilized for evaluating (Al-Amri, 2010; Goh & Burns, 2012; Namaziandost, Nasri, & Keshmirshekan, 2019; Sarwar, Alam, Hussain, Shah, & Jabeen, 2014). Holistic grading is a method of evaluating a composition based on its overall quality. Also known as global grading, single-impression scoring, and impressionistic grading. Developed by the Educational Testing Service, holistic grading is often used in large-scale assessments, such as college placement tests. Graders are expected to make judgments based on criteria that have been agreed upon before the start of an evaluation session. It will be contrasted with analytic grading. Holistic grading is useful as a time-saving approach, but it does not provide students with detailed feedback. The holistic scoring might be as well referred to as influential or global scale (Pan, 2016). The holistic approach is related to present a total scale, regarding the presentation entirely (Barylá, Shelley & Trainor, 2012; Griffith & Lim, 2012; Helvoort, 2010; Namaziandost & Nasri, 2019a; Schunn, Godley & DeMartino 2016). "An analytic or profile approach, on the other hand, tries to segregate out notable properties of execution and to assess every one exclusively and freely on its own subscale; the analytic approach thus therefore concentrates consideration on discrete characteristics of execution, normally mixing scores on the detached subscales to generate an overall score for speaking, and sometimes reporting the sub-scores too to give a more extravagant and wealthy dimension of source information, which can be beneficial for diagnostic objectives to manage future instructing/learning goals" (Taylor & Galaczi, 2011, p. 177).

As a result, some specific scales are applied in analytic rubrics (Allen & Tanner, 2006; Namaziandost & Nasri, 2019b). It is obvious that holistic scoring takes shorter time and is less complicated compared to the analytic approach. Even though, the analytic scoring provides plentiful information concerning learner's language proficiency (Kondo-Brown, 2002; Namaziandost, Rahimi Esfahani, & Ahmadi, 2019). Furthermore, grading accuracy is extended as graders' thought is appealed to the specific scale of language presentation (Luoma, 2004;

Namaziandost & Nasri, 2019c). Notwithstanding that analytical and global methods for scoring alter theoretically, they continuously overlap to somewhat (Taylor & Galaczi, 2011; Ziafar & Namaziandost, 2019).

Analytic Scoring

In examining speaking, analytic approach surveys various features of exam autonomously, scoring every belonging diversely (Richards & Schmidt, 2013). Applying analytical scoring inside the evaluation of spoken performance generates diverse benefits. Tuan (2012) asserts that it proposes efficient distinctive information on speaking ability of examinee, providing abundant wit into the learner's weaknesses and strengths. Jonsson and Svingby (2007) declare that it is besides the firmness of scoring amongst assignments, learners and diverse graders that is extended. In addition, applying analytical scoring promotes the trustiness of evaluation (Dogan & Uluman, 2017; Kaba & Sengül, 2016; Namaziandost, Shatalebi, & Nasri, 2019; Nasri, Biria, & Karimi, 2018). Ultimately, Finson, Ormsbee, and Jensen (2011, p. 181) state that "analytic rubrics bolster a progressively objective and reliable evaluation of learner work". Extended identity and firmness actually come out of applying the evaluation of a bit high points of spoken exam. Though the analytic method to the evaluation of oral skill demonstrates diverse noticeable benefits, it in the same way has some deficiencies. It takes more time and is troublesome since examiners need to present discrete scores for diverse sections of applicant's performance (Aleksandrak, 2011; Azadi, Biria, & Nasri, 2018; Hashemifardnia, Namaziandost, & Rahimi Esfahani, 2018a; Nasri & Biria, 2017; Saritha, 2016; Shatrova, Mullings, Blažejová, & Üstünel, 2017). Also, examiners have to be educated so as to trusty determine among various features and constituents of performance on the subject of how they are identified in the rubrics (Vafaei & Yaghmaeyan, 2015). Other deficiency is corona influence that alludes to the grading inside a scale might impact on the grading in other scale (Hashemifardnia, Namaziandost, & Sepehri, 2018; Hosseini, Nasri, & Afghari, 2017; Myford & Wolfe, 2003). Ultimately, Llach (2011, p. 57) expresses that "one of the major drawbacks of analytic scoring is the hardness in giving obvious and unequivocal definitions for each descriptor". Nevertheless, in spite of the fact that analytical scoring has some deficiencies, its values emerge to prevail and dominate the deficiencies, and following this kind of scoring inside the evaluation of performing speaking might be considered evenly appropriate.

Analytic Scoring Scale

To the degree that the explicit parts in analytic rubrics are considered, Pan (2016) explains that dimensionality for the evaluating the spoken skill might, utter; integrate fluency, authenticity and vocabulary. The Council of Europe (2001) includes the supplementary constituents of oral language: authenticity, fluency, range, adherence and interaction. As stated by Davies (1999) as usual applied classes in speaking exams are fluency, authenticity, pronunciation or comprehensibility and appropriateness. On the other part, Gondová (2014, p. 162) clarifies that "the accompanying criteria are regularly utilized: appropriateness, organization of ideas, fluency, grammatical accuracy and the range of grammatical structures, the range of vocabulary and its accuracy, content, pronunciation and intonation, and interaction" (Metruk, 2018; Hashemifardnia, Namaziandost, & Shafiee, 2018; Mirshekaran, Namaziandost, & Nazari, 2018). The scales of analytical evaluation in Cambridge English First certificate composed of pronunciation, vocabulary, grammar, speech management and interactive communication (Cambridge English: Understanding Results Guide, 2014). Tuan (2012) expresses that "based on the objective of the assessment, speaking performance might be evaluated on such criteria as content, organization, cohesion, register, vocabulary, grammar, or mechanics" (p. 673).

Amount of Scale

It is apparent that the option of particular classes has to arise out of the aim of evaluation. Nevertheless, examiners ought to be aware of the classes' quantity they apply as they estimate speaking. Normally their amount emerged between three and seven (Ruammai, 2014). In another view, Finson, Ormsbee, and Jensen (2011) state that three to six classes are applied on the whole. Nonetheless some questions are emerged regarding the utmost number of scales. "Received wisdom is that more than 4 or 5 classifications begins to cause cognitive overload and that 7 classifications are psychologically an upper bound" (Council of Europe 2001, p. 193). Analogously, Green (2014), Razali and Isra (2016), and Thornbury (2005) assert that four to five scales assume to be the most remarkable logical number concerning evaluating oral skill, as Luoma (2004) regards five to six classes to be the utmost. It seems to be rational to admit that it is next to incredible for examiners to focus on superior amount of scale than five or six, and direct rational and trusty evaluation simultaneously. "However, prior researches have not given sufficient

experimental proofs to help the designation of ideal number of criteria inside rating scales" (Chen, 2016, p. 52).

The connection between the analytic and holistic scoring of Iranian university EFL trainees' English oral skill was surveyed in current study. The members – the third- grade students of the study plan Teaching English Language took part six quarters of the English Language program, that was educated centered the Communicative Language Teaching (CLT) rules. In final exam, the trainees took an oral exam at C1 level based on Common European Framework of Reference for Languages (CEFR) within interview structure between speaking addressee and an applicant. Both kinds of analytic and holistic scoring were used. The analytic scoring was performed by an examiner, while the holistic one was performed by the addressee.

Four analytic scales were context and construction, vocabulary, pronunciation and grammar. The members were able to achieve the lowest of one and the utmost of five points in every class according to the exponents for every point that calculated for the whole 20 points.

The context and construction class composed of connection of responds to components, appropriate formation of long and short speeches, and answering to the components so as the communicative aim was conducted.

The principal point of the section of pronunciation was integrated in direction of conceivability beside the well likely enunciation of singular phonemes and proper use of stress and intonation. Due to the fact that L2 speakers' English speeches normally demonstrate abnormal phonetic discoveries regarding to their L1 (Bilá, 2010; Namaziandost, Sabzevari, & Hashemifardnia, 2018), insignificant and worthless components of L1 accent in the members' formation were not punished.

The grammar and vocabulary criteria estimated range, as well as precision (Namaziandost, Hosseini, & Utomo, 2020). As far as the vocabulary category as such is involved, Topkaraoğlu and Dilman (2014) demonstrate that the number of words an L2 learner knows does not appear to be adequate; the participants additionally require to have considerable amount of information about the words they have procured if they desire to become effectual and efficient users of a foreign language. Finally, attention was also devoted to grammar. Similar to vocabulary, both grammatical range and accuracy were inspected.

Concerning the scale of holistic, the trainees were able to achieve the

lowest of one and the utmost of five points according to the exponents for every point. Hence, the members could achieve whole 25 points for the whole evaluation (analytic scoring + holistic scoring). For example, a trainee achieved 4 points for context and construction, 3 points to pronunciation, 4 points to vocabulary and 3 points to grammar from the examiner, and the speaking addressee presented them 3 points. Thoroughly, they marked (4 + 3 + 4 + 3 + 3) 17 points out of 25, which constitutes 68%. Thenceforth the succeeding questions were created.

1. Which scores carry out the topics attain in the four classes of analytic scoring?
2. What is the mean score regarding the holistic scoring?
3. What is the mean score regarding the analytic scoring?
4. What is the diversity between analytic and holistic scoring? Is the diversity statistically particular?

RESEARCH METHODOLOGY

Members

Seventy second-grade university EFL trainees were members of the study who were learning English Language Teaching at state university of Malayer, Iran. The members in the study aged from 24 to 28, the number of males and females were (n=30) and (n=40) respectively. The speaking addressee and examiner were two Iranian Ph.D. holders of TEFL. They had approximately six years of experience in evaluating the oral skill as the evaluation was performed, and the examiner had conducted two quarters of evaluating English language program as a section of his master and Ph.D. studies.

Procedures and instruments

The members were given a subject by accident which they were required to have an interview with the speaking addressee. The addressee questioned idea-based open questions, which were in the range of universal knowledge of the topics; hence the evaluation procedure was not contrarily influenced by examining knowledge more than oral skills. The examiner was nearby in order not to disturb or affect the members. He was noticing due to create his evaluation as trusty as potential. The exam took near 15 minutes. After a while, a member was inquired to stand out of the room, hence the speaking addressee and examiner were able to give points to the members for their execution. Whereas the whole

score was calculated, the member was return back to the room in order to argue how they conduct the speaking exam. Every member was assigned a useful feedback on how they conduct each class.

Results

The grades of analytic scoring with every classes' scores (context and construction, vocabulary, grammar, pronunciation) are displayed in Table 1. The Table as well contains the average values of all topics' execution in four classes. The information illustrate that the members were extremely rewarding in the class context and construction (3.928), and achieved the lowest scores in class of grammar (2.514). The vocabulary and pronunciation sections signify the scores 3.471 and 3.271 specifically.

Table 1:

Analytic scoring grades

| Member | Context | Pronunciation | Vocabulary | Grammar |
|--------|---------|---------------|------------|---------|
| 1 | 4 | 3 | 3 | 4 |
| 2 | 3 | 3 | 3 | 1 |
| 3 | 4 | 3 | 3 | 1 |
| 4 | 3 | 3 | 3 | 2 |
| 5 | 2 | 3 | 2 | 2 |
| 6 | 3 | 2 | 2 | 3 |
| 7 | 4 | 2 | 1 | 3 |
| 8 | 3 | 2 | 2 | 3 |
| 9 | 3 | 5 | 3 | 3 |
| 10 | 4 | 5 | 3 | 4 |
| 11 | 4 | 4 | 4 | 4 |
| 12 | 3 | 2 | 4 | 5 |
| 13 | 3 | 2 | 4 | 3 |
| 14 | 3 | 3 | 4 | 3 |
| 15 | 3 | 3 | 4 | 2 |
| 16 | 5 | 4 | 2 | 2 |
| 17 | 3 | 4 | 2 | 2 |
| 18 | 4 | 4 | 2 | 2 |
| 19 | 5 | 2 | 1 | 1 |
| 20 | 2 | 3 | 3 | 1 |
| 21 | 5 | 3 | 5 | 1 |
| 22 | 5 | 5 | 5 | 2 |
| 23 | 5 | 5 | 3 | 2 |
| 24 | 5 | 5 | 5 | 2 |
| 25 | 2 | 5 | 4 | 2 |
| 26 | 4 | 5 | 4 | 3 |
| 27 | 5 | 4 | 3 | 3 |
| 28 | 3 | 4 | 3 | 3 |
| 29 | 5 | 4 | 4 | 3 |
| 30 | 5 | 4 | 2 | 4 |

| | | | | |
|------|-------|-------|-------|-------|
| 31 | 4 | 2 | 4 | 5 |
| 32 | 4 | 2 | 4 | 3 |
| 33 | 5 | 2 | 4 | 3 |
| 34 | 5 | 3 | 4 | 2 |
| 35 | 5 | 3 | 3 | 2 |
| 36 | 5 | 4 | 4 | 2 |
| 37 | 4 | 4 | 3 | 2 |
| 38 | 5 | 4 | 4 | 4 |
| 39 | 4 | 4 | 3 | 4 |
| 40 | 4 | 3 | 4 | 3 |
| 41 | 5 | 4 | 3 | 3 |
| 42 | 5 | 5 | 5 | 1 |
| 43 | 3 | 4 | 3 | 2 |
| 44 | 4 | 4 | 5 | 3 |
| 45 | 4 | 3 | 2 | 1 |
| 46 | 5 | 4 | 3 | 4 |
| 47 | 5 | 3 | 3 | 5 |
| 48 | 5 | 4 | 1 | 2 |
| 49 | 5 | 2 | 1 | 2 |
| 50 | 5 | 2 | 3 | 2 |
| 51 | 4 | 4 | 3 | 3 |
| 52 | 4 | 3 | 5 | 3 |
| 53 | 2 | 5 | 4 | 4 |
| 54 | 2 | 5 | 4 | 2 |
| 55 | 3 | 4 | 4 | 5 |
| 56 | 3 | 3 | 2 | 3 |
| 57 | 3 | 5 | 2 | 2 |
| 58 | 4 | 4 | 3 | 2 |
| 59 | 5 | 4 | 5 | 2 |
| 60 | 5 | 3 | 5 | 2 |
| 61 | 4 | 3 | 3 | 1 |
| 62 | 4 | 3 | 3 | 1 |
| 63 | 3 | 3 | 4 | 1 |
| 64 | 3 | 4 | 4 | 2 |
| 65 | 2 | 2 | 4 | 3 |
| 66 | 2 | 4 | 3 | 1 |
| 67 | 5 | 3 | 3 | 1 |
| 68 | 5 | 3 | 2 | 1 |
| 69 | 5 | 3 | 3 | 2 |
| 70 | 4 | 3 | 4 | 3 |
| Mean | 3.928 | 3.471 | 3.271 | 2.514 |

The context and construction part were the lowest debatable one among the four parts. The members were compensated for not joining to the point, or the time the questions were not responded and the speeches were either irrelevant, not fluent or of an inappropriate extent. The class of pronunciation consisted of both super segmental and segmental fallacies. The parts commonly implied the

substitution of English phonemes, specifically ones that were not in topics' L1, to Iranian tones. "Both teachers and learners require to reminisce that substituting some sounds for others hampers communication and mostly causes a menace to intelligibility" (Metruk, 2017, p. 15). The uppermost common fault in the prosodic aspects was stress of word. Concerning the class of grammar and vocabulary, the trainees encounter to significant troubles and difficulties with the extent of lexis, and accomplished even great difficulties with the extent of grammar structures.

Table 2 demonstrates the mean of analytic scoring grade for every member. For example, on the condition that a member received 5 points for context and construction, grade 3 to pronunciation, 4 to vocabulary and grade 3 to grammar, the mean grade for analytic scoring is 3.5 ($5 + 3 + 4 + 3 = 15$, and the grade was divided by the classes' number: $15 \div 4 = 3.75$). The holistic scoring mean of all members was 3.628 which were also displayed in Table 2 (Appendix B), whereas the mean value of analytic scoring was 3.396 for all members. However, the diversity of analytic and holistic scoring is just 0.232 ($3.628 - 3.396 = 0.232$), the p-value was computed at 0.002 for the statistical significance level. It shows a statistical considerable diversity between analytic and holistic scoring ($p < 0.05$).

Thus, the research results disclose that the method of analytic scoring illustrated more accurate and trustier trend of evaluating the oral skill compared to the method of holistic scoring. Furthermore, the members supplied with exact feedback on how rewarding they were in every class as the examiner noticed along the exam. The analytic scoring in addition showed distinctive information thus the trainers realized what fields the EFL trainees need to notice more subsequently.

Table 1:

Comparison of analytic and holistic scoring

| Member | Analytic scoring mean | Holistic scoring |
|--------|-----------------------|------------------|
| 1 | 2.75 | 3 |
| 2 | 2.75 | 4 |
| 3 | 2 | 3 |
| 4 | 2 | 5 |
| 5 | 3.5 | 3 |
| 6 | 2 | 2 |
| 7 | 2.5 | 3 |
| 8 | 2.5 | 5 |
| 9 | 4 | 5 |
| 10 | 4 | 3 |
| 11 | 3.75 | 4 |

| | | |
|----|------|---|
| 12 | 2 | 2 |
| 13 | 2 | 5 |
| 14 | 3.25 | 2 |
| 15 | 3 | 2 |
| 16 | 2.25 | 4 |
| 17 | 4 | 3 |
| 18 | 3.75 | 5 |
| 19 | 2 | 4 |
| 20 | 3.75 | 3 |
| 21 | 3 | 4 |
| 22 | 3.75 | 5 |
| 23 | 5 | 2 |
| 24 | 5 | 4 |
| 25 | 3.75 | 3 |
| 26 | 5 | 2 |
| 27 | 4 | 5 |
| 28 | 3.25 | 4 |
| 29 | 4 | 3 |
| 30 | 4 | 3 |
| 31 | 2 | 3 |
| 32 | 3.25 | 2 |
| 33 | 3.25 | 3 |
| 34 | 3.25 | 4 |
| 35 | 3 | 4 |
| 36 | 4 | 5 |
| 37 | 3.5 | 3 |
| 38 | 3.75 | 4 |
| 39 | 4.5 | 4 |
| 40 | 3 | 5 |
| 41 | 4.75 | 4 |
| 42 | 4.75 | 4 |
| 43 | 4.75 | 4 |
| 44 | 4.75 | 2 |
| 45 | 3 | 3 |
| 46 | 2.25 | 3 |
| 47 | 3.25 | 4 |
| 48 | 2.25 | 3 |
| 49 | 2.75 | 5 |
| 50 | 2 | 3 |
| 51 | 4.25 | 4 |
| 52 | 3 | 3 |
| 53 | 4.25 | 5 |
| 54 | 4.25 | 4 |
| 55 | 4.25 | 4 |
| 56 | 3.25 | 2 |
| 57 | 2.5 | 4 |
| 58 | 4 | 4 |
| 59 | 4.25 | 3 |
| 60 | 3 | 4 |
| 61 | 3.25 | 3 |

| | | |
|------|-------|-------|
| 62 | 2 | 5 |
| 63 | 3 | 3 |
| 64 | 3.75 | 4 |
| 65 | 3.75 | 5 |
| 66 | 3.75 | 4 |
| 67 | 3.75 | 4 |
| 68 | 3 | 5 |
| 69 | 3.5 | 3 |
| 70 | 4.5 | 5 |
| Mean | 3.396 | 3.628 |

DISCUSSION AND CONCLUSION

This research aimed to survey the analytic and holistic trend of evaluating oral skills in a superior-education context. The outcomes show that the members in four classes – context and construction, vocabulary, pronunciation and grammar gained 3.928, 3.271, 3.471 and 2.514 points respectively. Despite the CLT must be necessary method for TEFL. It sounds that L2 trainees face to troubles as they require applying C1/B2 level words beside more complicated and fake structures of grammar in their speeches. It might be the conclusion of using the Grammar Translation Method to some extent in Iranian educational system. The trainees might realize the C1/B2 words; Of course, they are not prepared to apply them during their speech. Consequently, giving EFL trainees adequate space to exercise speaking and pursuing the CLT rules would be beneficial.

As well the findings of the research clarify that the mean score of analytic and holistic trends of scoring respectively was 3.396 and 3.628. The p-value was calculated at 0.002; thus, a statistically considerable diversity found between the analytic and holistic trends of scoring ($p < 0.05$). It does not convey that one scoring method is trustier than another while the mentality of the examiner and the speaking addressee might have acted its role. Albeit using both trends of scoring in the evaluation procedure would be observed as appropriate and costly as the two methods seem to augment one another. Likewise, analytic scoring empowered the subjects to be equipped by a comprehensive feedback on their execution in particular classes. At last, the outcomes proposed useful distinctive information thus both the trainers and EFL superior-education trainees realize what fields they must concentrate extensively.

This research undergoes some restrictions. First, there were just one speaking

addressee and one examiner and their mental realization and analysis of a member oral presentation may have influenced on the evaluation procedure. Though it must be stated that the estimating oral skill is an exceedingly mental procedure and there are infinite elements which influence on examiner's assessment (Jankowska & Zielińska, 2015; Namaziandost, & Nasri, Rahimi Esfahani, & Keshmirshekan 2019). It is hence propounded that next researches apply an extreme number of examiners due to provide sufficient statistical capability for estimation of the connection between the analytic and holistic trends of scoring.

Similarly, an enormous sample of members can be taken part in next studies too. In addition, the explanation of bands in the scales of analytical scoring might have acted its role in evaluation procedure. Further a mental explanation may have influenced on the evaluation procedure. Nonetheless, it must be emphasized that it is fairly rigid duty to recommend explicit and absolute descriptions for the exponents (Llach, 2011; Namaziandost, Neisi, Kheryadi, & Nasri, 2019). It seems rational to assume that the mentality level can be decreased by going through a proper training and by achieving periods of experience, and then the evaluation can develop as exact, trusty and concrete as possible. At last contrasting diversity between scores of female and male by evaluating oral skill in next researches may be engaging.

It can be concluded that integrating these two holistic and analytic scoring may be considered as somewhat applicable option as it occurs in evaluation of speaking skills. Both trends of scoring have their benefits and deficiencies and applying them may direct to a more concrete scoring.

REFERENCES

- Abedi, P., Keshmirshekan, M. H., & Namaziandost, E. (2019). The comparative effect of flipped classroom instruction versus traditional instruction on Iranian intermediate EFL learners' English composition writing. *Journal of Applied Linguistics and Language Research*, 6(4), 43-56.
- Al-Amri, M. (2010). Direct spoken English testing is still a real challenge to be worth bothering about. *English Language Teaching*, 3 (1), 113-117.
- Aleksandrzak, M. (2011). Problems and challenges in teaching and learning speaking at advanced level. *Glottodidactica*, 37, 37-48.
- Allen, D. & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *Life Sciences Education*, 5 (3), 197-203.

- Anderson, J. (2015). *A guide to the practice of English language teaching for teachers and trainee teachers*. Nairobi: East African Educational Publishers Ltd.
- Azadi, G., Biria, R., & Nasri, M. (2018). Operationalising the Concept of Mediation in L2 Teacher Education. *Journal of Language Teaching and Research*, 9(1), 132-140.
- Baryla, E., Shelley, G., & Trainor, W. (2012). Transforming rubrics using factor analysis: Practical assessment. *Practical Assessment Research & Evaluation*, 17 (4).
- Bilá, M. (2010). Perception and production of a second language and the concept of a foreign accent. In S. Pokrivčáková et al. (Eds.) *Modernization of teaching foreign languages: CLIL, inclusive and intercultural education*, pp. 123-143. Brno: Masaryk University.
- Cambridge English: Understanding Results Guide. (2014). http://www.gml.cz/prof/zajickova/Cambridge%20exams_information/Understanding%20results%20guide.pdf
- Chen, G. (2016). Developing a Model of analytic rating scales to assess college students' 12 Chinese oral performance. *International Journal of Language Testing*, 6 (2), 50-71.
- Chuang, Y. (2009). Foreign language speaking assessment: Taiwanese college English teachers' scoring performance in the holistic and analytic rating methods. *The Asian EFL Journal*, 11 (1), 150-173.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.
- Davies, A. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Dogan, C. & Uluman. M. (2017). A comparison of rubrics and graded category rating scales with various methods regarding raters' reliability. *Educational Sciences: Theory & Practice*, 17 (2), 631-651. doi:10.12738/estp.2017.2.0321.
- Finson, K., Ormsbee, C., & Jensen, M. (2011). *Differentiating science instruction and assessment for learners with special needs, k-8*. Thousand Oaks: Corwin, A SAGE Company.
- Goh, C. & Burns, A. (2012). *Teaching speaking: A holistic approach*. New York: Cambridge University Press.
- Göktürk, N. (2016). Examining the Effectiveness of Digital Video recordings on Oral Performance of EFL Learners. *Teaching English with Technology*, 16 (2), 71-96.

- Gondová, D. (2014). Taking first steps in teaching English: Assessing learners. Žilina: EDIS.
- Green, A. (2014). Exploring language assessment and testing: Language in action. New York: Routledge.
- Griffith, W. & Lim, H. (2012). Performance-based assessment: rubrics, web 2.0 tools and language competencies. *Mextesol Journal*, 36 (1).
- Hashemifardnia, A., Namaziandost, E., & Rahimi Esfahani, F. (2018a). The Effect of Teaching Picture-books on Elementary EFL Learners' Vocabulary Learning. *JELTL (Journal of English Language Teaching and Linguistics)*, 3(3), 247-258.
- Hashemifardnia, A., Namaziandost, E., & Sepehri, M. (2018). The effectiveness of giving grade, corrective feedback, and corrective feedback-plus-giving grade on grammatical accuracy. *International Journal of Research Studies in Language Learning*, 8 (1), 15-27.
- Hashemifardnia, A., Namaziandost, E., Shafiee, S. (2018). The Effect of Implementing Flipped Classrooms on Iranian Junior High School Students' Reading Comprehension. *Theory and Practice in Language Studies*, 8(6), 665-673.
- Helvoort, J. (2010). A scoring rubric for performance assessment of information literacy in Dutch higher education. *Journal of Information Literacy*, 4 (1), 22-39.
- Hosseini, E. Z., Nasri, M., & Afghari, A. (2017). Looking beyond teachers' classroom behavior: novice and experienced EFL teachers' practice of pedagogical Knowledge to Improve Learners' Motivational Strategies. *Journal of Applied Linguistics and Language Research*, 4(8), 183-200.
- Jankowska, A. & Zielińska, U. (2015). Designing a self-assessment instrument for developing the speaking skill at the advanced level. In M. Pawlak and E. WaniekKlimczak (Eds.) *Issues in teaching, learning and testing speaking in a second language*. Heidelberg: Springer.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.
- Kaba, Y. & Sengül, S. (2016). Developing the rubric for evaluation problem posing (REPP). *International Online Journal of Educational Sciences*, 8 (1), 8-25.
- Khamkhien, A. (2010). Teaching English speaking and English-speaking tests in the Thai context: a reflection from Thai perspective. *English Language Teaching*, 3 (1), 184-190.
- Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing*, 19 (1), 3-31.

- Llach, M. (2011). *Lexical errors and accuracy in foreign language writing*. New York: Multilingual Matters.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Metruk, R. (2017). Pronunciation of English dental fricatives by Slovak University EFL Students. *International Journal of English Linguistics*, 7 (3), 11-16.
- Mirshekaran, R., Namaziandost, E., & Nazari, M. (2018). The Effects of Topic Interest and L2 Proficiency on Writing Skill among Iranian EFL Learners. *Journal of Language Teaching and Research*, 9(6), 1270-1276.
- Murugaiah, P. (2016). Pecha Kucha style PowerPoint presentation: An innovative CALL approach to developing oral presentation skills of tertiary students. *Teaching English with Technology*, 16 (1), 88-104.
- Myford, C. & Wolfe, E. (2003). Detecting and measuring rater effects using many facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4 (4), 386-422.
- Namaziandost E., & Nasri, M. (2019a). A meticulous look at Long's (1981) interaction hypothesis: does it have any effect on speaking skill? *Journal of Applied Linguistics and Language Research*, 6(2), 218-230.
- Namaziandost E., & Nasri, M. (2019b). The impact of social media on EFL learners' speaking skill: A survey study involving EFL teachers and students. *Journal of Applied Linguistics and Language Research*, 6(3), 199-215.
- Namaziandost E., & Nasri, M. (2019c). Innovative practices in L2 writing materials in the EFL classroom: Effect on writing enhancement and attitude to English course. *Asian Journal of Research in Social Sciences and Humanities*, 9(9), 1-12.
- Namaziandost, E., & Nasri, M., Rahimi Esfahani, F., & Keshmirshekan M. H. (2019). The impacts of spaced and massed distribution instruction on EFL learners' vocabulary learning. *Cogent Education*, 6: 1661131. <https://doi.org/10.1080/2331186X.2019.1661131>
- Namaziandost, E., Shatalebi, V., & Nasri, M. (2019). The impact of cooperative learning on developing speaking ability and motivation toward learning English. *Journal of Language and Education*, 5(3), 83-101. <https://doi.org/10.17323/jle.2019.9809>.
- Namaziandost, E., Nasri, M., & Keshmirshekan, M. H. (2019). Cohesive conjunctions in applied linguistics research articles among Iranian and non-Iranian researchers: A comparative corpus-based study. *Journal of English Language Studies*, 4(2), 101 – 119.
- Namaziandost, E., Neisi, L., Kheryadi, & Nasri, M. (2019). Enhancing oral proficiency through cooperative learning among intermediate EFL learners: English learning motivation in focus. *Cogent Education*, 6(1), 1-15.

<https://doi.org/10.1080/2331186X.2019.1683933>

- Namaziandost, E., Sabzevari, A., & Hashemifardnia, A. (2018). The effect of cultural materials on listening comprehension among Iranian upper-intermediate EFL learners: In reference to gender. *Cogent Education*, 5(1), 1-27. <https://doi.org/10.1080/2331186X.2018.1560601>.
- Namaziandost, E., Rahimi Esfahani, F., & Ahmadi, S. (2019). Varying levels of difficulty in L2 reading materials in the EFL classroom: Impact on comprehension and motivation. *Cogent Education*, 6(1), 1-9. <https://doi.org/10.1080/2331186X.2019.1615740>.
- Namaziandost, E., Hosseini, E., & Utomo, D. W. (2020). A comparative effect of high involvement load versus lack of involvement load on vocabulary learning among Iranian sophomore EFL learners. *Cogent Arts & Humanities*, 7(1), 1715525. <https://doi.org/10.1080/23311983.2020.1715525>
- Nasri, M. & Biria, R. (2017). Integrating multiple and focused strategies for improving reading comprehension and L2 lexical development of Iranian intermediate EFL learners. *International Journal of Applied Linguistics & English Literature*, 6(1), 311-321.
- Nasri, M., Biria, R., & Karimi, M. (2018). Projecting Gender Identity in Argumentative Written Discourse. *International Journal of Applied Linguistics & English Literature*, 7(3), 201-205.
- Nasri, M., Namaziandost, E., & Akbari, S. (2019). Impact of pictorial cues on speaking fluency and accuracy among Iranian pre-intermediate EF learners. *International Journal of English Language and Literature Studies*, 8(3), 99-109
- O' Sullivan, B. (2012). Assessing speaking. In C. Coombe, P. Davidson, B. O' Sullivan, and S. Stoyhoff (Eds.) *the Cambridge guide to second language assessment*, pp. 234- 246. New York: Cambridge University Press.
- Pan, M. (2016). Nonverbal delivery in speaking assessment. from an argument to a rating scale formulation and validation. Singapore: Springer.
- Pokrivčáková, S. (2010). *Modern teacher of English*. Nitra: ASPA. Razali, K. & Isra, M. (2016). Male and female teachers' roles in assessment of speaking skill: Gender equality: *International Journal of Child and Gender Studies*, 2 (1), 1-10.
- Richards, J. & Schmidt, R. (2013). *Longman dictionary of language teaching and applied linguistics* (4th ed.). New York: Routledge.
- Ruammai, P. (2014). Constructing scoring instrument for writing assessment and fostering critical thinking. In H. Lee (Ed.) *The international conference on language and communication innovative inquiries and emerging paradigms in language, media and communication*, pp. 127-137.

- Saritha, K. (2016). Rubric for English language teaching research. *Research Journal of English Language and Literature*, 4 (2), 725-731.
- Sarwar, M., Alam, M., Hussain, S., Shah, A., & Jabeen, M. (2014). Assessing English speaking skills of prospective teachers at entry and graduation level in teacher education program. *Language Testing in Asia*, 4 (5).
- Schunn, C., Godley, A., & DeMartino, S. (2016). The reliability and validity of peer review of writing in high school English classes. *Journal of Adolescent & Adult Literacy*, 60 (1), 13-23. doi:10.1002/jaal.525
- Shatrova, Z., Mullings, R., Blažejová, S., & Üstünel, E. (2017). English speaking assessment: developing a speaking test for students in a preparatory school. *International Journal of English Language Teaching*, 5 (3), 27-40.
- Taylor, L. & Galaczi, E. (2011). Scoring validity. In L. Taylor (Ed.), M. Milanovic & C. Weir (Series Eds.) *Studies in language testing 30. examining speaking. research and practice in assessing second language speaking*, pp. 171-233. Cambridge: Cambridge University Press.
- Thornbury, S. (2005). *How to teach speaking*. Harlow: Pearson Education Limited.
- Topkaraoğlu, M. & Dilman, H. (2014). Effects of studying vocabulary enhancement activities on students' vocabulary production levels. *Procedia - Social and Behavioral Sciences*, 152, 931-936. <https://doi.org/10.1016/j.sbspro.2014.09.345>
- Tuan, L. (2012). Teaching and assessing speaking performance through analytic scoring approach. *Theory and Practice in Language Studies*, 2 (4), 673-679. doi:10.4304/tpls.2.4.673-679. <https://doi.org/10.4304/tpls.2.4.673-679>
- Ur, P. (2012). *A course in English language teaching*. Cambridge: Cambridge University Press.
- Vafee, P. & Yaghmaeyan, B. (2015). Providing evidence for the generalizability of a speaking placement test scores. *Iranian Journal of Language Testing*, 5 (2), 78-95.
- Ziafar M., & Namaziandost, E. (2019). Linguistics, SLA and lexicon as the unit of language. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 2(5), 245-250