

EVALUATING ORAL PROFICIENCY SKILL THROUGH ANALYTICS AND HOLISTIC WAYS OF SCORING

Ehsan Namaziandost^{1*}, Reza Banari², Shahrzad Momtaz³

¹PhD Candidate in TEFL, Department of English, Faculty of Humanities, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran, ²PhD in TEFL, Department of English, Baghmalek Branch, Islamic Azad University, Baghmalek, Iran, ³Department of English, Faculty of Humanities, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran.
Email: ^{1*}e.namazi75@yahoo.com, ²rezabanari2010@yahoo.com, ³shahrzadmomtaz@yahoo.com

Article History: Received on 28th July 2019, Revised on 30th August 2019, Published on 05th October 2019

Abstract

Purpose: Evaluating speaking skill is an exceedingly difficult and intricate subject. Two methods of testing oral proficiency are usually used: holistic and analytic scoring. To this end, this study aimed to compare analytic and holistic techniques for scoring in evaluating oral proficiency skills.

Methodology: The participants of this study were 70 second-grade university EFL students who were studying English Language Teaching at a university in Iran. The participants in this research were both male (n=40) and female (n=30) with the age range from 24 to 25.

Main Findings: The results showed a statistically considerable diversity between analytic and holistic methods of evaluation considering that the p-value was estimated at 0.002 ($P < 0.05$).

Applications: The findings confirmed that employing these two scoring techniques in the procedure of evaluation may be considered proper as they seem to complement each other, one augments the other and lead to more inclusive evaluation.

Novelty/Originality: This study is new since it is the first time in an Iranian university context; oral proficiency skill was evaluated through holistic and analytic scoring.

Keywords: oral skill evaluation, analytic scoring, holistic scoring, teaching English as a foreign language, inclusive evaluation.

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: “[w]ith 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 (Keshmirshakan, Namaziandost, & Pournorouz, 2019; Murugaiah, 2016). Although concurrently, speaking might be regarded as the supreme disturbing skill to require as a language has to be generated promptly and random, that requires a significant deal of exercise (Anderson, 2015; Nasri, Namaziandost, & Akbari, 2019). Undoubtedly, it takes a long time and constant attempt for a foreign language learner to become skillful in speaking skills.

Concerning the evaluation of speaking skills, O’Sullivan (2012, p.234) affirms that “[i]t 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 influence examiners, evaluating speaking presentations seems to be one of the utmost rigid duties to the executive. As well, Luoma (2004) likewise asserts that speaking evaluation is provoking due to the fact that there are abundant elements that influence the conception of an examiner concerning how great an individual is able of speaking. In addition, examiners consider examining 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 presentations is a slightly rigid duty and requires a lot of features to be thought.

LITERATURE REVIEW

Approaches to 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, & Keshmirshakan, 2019; Sarwar, Alam, Hussain, Shah, & Jabeen, 2014). 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 ([Baryla, Shelley & Trainor, 2012](#); [Namaziandost&Nasri, 2019a](#)). 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.

As a result, some specific scales are applied in analytic rubrics ([Namaziandost&Nasri, 2019b](#)). It is obvious that holistic scoring takes a 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, Nasri, & Mirshekaran, 2018](#)). Furthermore, grading accuracy is extended as graders' thought is appealed to the specific scale of language presentation ([Namaziandost&Nasri, 2019c](#)). Notwithstanding that analytical and global methods for scoring alter theoretically, they continuously overlap to somewhat.

Analytic scoring

In examining speaking, the analytic approach surveys various features of the exam autonomously, scoring every belonging diversely. 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 beside the firmness of scoring amongst assignments, learners and diverse graders that are extended. In addition, applying analytical scoring promotes the trustiness of evaluation ([Namaziandost, Abedi, & Nasri, 2019](#); [Nasri, Biria, & Karimi, 2018](#)). Ultimately, [Finson, Ormsbee, and Jensen \(2011, p. 181\)](#) state that "[a]nalytic 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 is 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 ([Azadi, Biria, & Nasri, 2018](#); [Hashemifardnia, Namaziandost, & Rahimi Esfahani, 2018a](#); [Nasri&Biria, 2017](#); [Saritha, 2016](#)). 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. Another deficiency is corona influence that alludes to the grading inside a scale might impact on the grading in other scales ([Hashemifardnia, Namaziandost, & Sepehri, 2018](#); [Hosseini, Nasri, & Afghari, 2017](#)). Ultimately, [Llach \(2011, p. 57\)](#) expresses that "[o]ne of the major drawbacks of analytic scoring is the hardness is 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 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 usually 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 to the whole. Nonetheless, some questions have emerged regarding the utmost number of scales. "Received wisdom is that more than 4 or 5 classifications begin 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](#)).

Research history

The connection between the analytic and holistic scoring of Iranian university EFL trainees’ English oral skills was surveyed in the 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 response to components, the 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 into the 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 their L1 ([Bilá, 2010](#); [Namaziandost, Nasri, & Rahimi Esfahani, 2019b](#); [Ziafar&Namaziandost, 2019](#)), insignificant and worthless components of L1 accent in the members’ formation were not punished.

The vocabulary and grammar scale evaluated range, beyond accuracy. To the extent that the vocabulary class is implied, [Topkaraoğlu and Dilman \(2014\)](#) show that the quantity of words an L2 trainee realizes does not seem to be sufficient; the members as well need to possess noticeable amount of information regarding the words they have acquired on the condition that they want to be efficient and virtuous trainees of a foreign language. Ultimately, grammar was noticed. Alike the vocabulary, both accuracy, and grammar were investigated.

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 a 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%. Therefore, 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 holistic scoring?
3. What is the mean score regarding the analytic scoring?
4. What is the diversity between analytic and holistic scoring? Is diversity statistically particular?

METHOD

Members

Seventy-second-grade university EFL trainees were members of the study who were learning English Language Teaching at the University of Ahvaz, Iran. The members in the study aged from 24 to 25, 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. They were given no time to being prepared. 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 creating 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 returned back to the room in order to argue how they conduct the speaking exam. Every member was assigned 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 (see Appendix A). The Table as well contains the average values of all topics' execution in four classes. The information illustrates that the members were extremely rewarding in the class context and construction (3.928), and achieved the lowest scores in the class of grammar (2.514). The vocabulary and pronunciation sections signify the scores 3.471 and 3.271 specifically.

The context and construction part was 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 the stress of word. Concerning the class of grammar and vocabulary, the trainees encounter significant troubles and difficulties with the extent of lexis and accomplished even great difficulties with the extent of grammar structures.

Table 2 (see Appendix B) demonstrates the mean of the 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 statistically 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.

DISCUSSION AND CONCLUSION

This research aimed to survey the analytic and holistic trend of evaluating oral skills in a superior-education context. Entirely 70 second-grade university trainees undertook an oral exam in the TEFL at Ahvaz University in Iran. The level of the exam was C1 and according to CEFR. Both analytic and holistic scoring trends were applied. 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 a necessary method for TEFL. It sounds that L2 trainees face troubles as they require applying C1/B2 level words besides 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; albeit 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 with 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 (Shakibaei, Shahamat, & Namaziandost, 2019).

This research undergoes some restrictions. First, there was just one speaking addressee and one examiner and their mental realization and analysis of a member oral presentation may have influenced the evaluation procedure. Though it must be stated that the estimating oral skill is an exceedingly mental procedure and there are infinite elements that influence on examiner's assessment (Jankowska&Zielińska, 2015; Namaziandost, & Nasri, Rahimi Esfahani, & Keshmirshakan 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 apart in the 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 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, Nasri, & Rahimi Esfahani, 2019a). It seems rational to

assume that the mentality level can be decreased by going through 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 females and males by evaluating oral skills in next researches may be engaging.

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

LIMITATION AND STUDY FORWARD

It is worth to mention that some limitations need to be taken into account about the present study. The first one was the small size of the research sample, so the findings must be generalized and used with caution. Moreover, this study included only participants that were 24-25 years old. So, the results cannot be generalized to the other age groups. Likewise, the role of attitudes, anxiety level and family situations of the participants were not controlled by the researcher. Another limitation was that the participated learners in the present study were second-grade university EFL students regarding language proficiency; the next studies are offered to include other levels- intermediate, advanced, and elementary learners. Furthermore, the current research was done in EFL context and examined oral performance and ESL context and other language skills were ignored in this study. Finally, the result of this study may be affected by the classroom situation and social factors. These factors have not been taken into account in the present study.

During conducting the present study some recommendations were formed in the researcher's mind. The first suggestion for the next studies is to include more participants to get richer findings. The second recommendation for the future studies is to work on other language proficiency levels-intermediate, upper-intermediate and advanced. The third suggestion for the upcoming researches is to investigate other language skills and sub-skills. The fourth suggestion for the next studies with a similar topic is to take the gender into account. Variables other than the ones used in this study can help the investigation of the same issue. For instance, the same experiment with male or female learners within the same age range is necessary to confirm the result of this study.

REFERENCES

1. Abedi, P., Keshmirshakan, 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. <http://www.jallr.com/index.php/JALLR/article/view/1036>
2. 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. https://pdfs.semanticscholar.org/c5a7/99_0d5eb6da7e2762e4c86717b61766823510.pdf. <https://doi.org/10.5539/elt.v3n1p113>
3. Anderson, J. (2015). *A guide to the practice of English language teaching for teachers and trainee teachers*. Nairobi: East African Educational Publishers Ltd.
4. 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. <http://www.academypublication.com/ojs/index.php/jltr/article/view/jltr0901132140>. <https://doi.org/10.17507/jltr.0901.17>
5. Baryla, E., Shelley, G., & Trainor, W. (2012). Transforming rubrics using factor analysis: Practical assessment. *Practical Assessment Research & Evaluation*, 17 (4), 1-17. <https://pareonline.net/getvn.asp?v=17&n=4>
6. 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.
7. Cambridge English: Understanding Results Guide. (2014). http://www.gml.cz/prof/zajickova/Cambri dge%20exams_information/Understanding%20results%20guide.pdf
8. 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. http://ijlte-1.ov2.com/content/ijlte_1_ov2_com/wp-content_138/uploads/2019/07/443-2016-6-2.pdf
9. 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. <https://www.asian-efl-journal.com/main-journals/foreign-language-speaking-assessment-chinese-taiwanese-college-english-teachers-scoring-performance-in-the-holistic-and-analytic-rating-methods/>
10. Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.
11. Davies, A. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
12. 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. <https://doi.org/10.4135/9781483387529>

13. Goh, C. & Burns, A. (2012). Teaching speaking: A holistic approach. New York: Cambridge University Press.
14. Göktürk, N. (2016). Examining the Effectiveness of Digital Video recordings on Oral Performance of EFL Learners. *Teaching English with Technology*, 16 (2), 71-96. <https://eric.ed.gov/?id=EJ1135917>
15. Gondová, D. (2014). *Taking first steps in teaching English: Assessing learners*. Žilina: EDIS.
16. Green, A. (2014). *Exploring language assessment and testing: Language in action*. New York: Routledge. <https://doi.org/10.4324/9781315889627>
17. 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. <https://jeltl.org/index.php/jeltl/article/view/151>. <https://doi.org/10.21462/jeltl.v3i3.151>
18. 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. <http://consortiacademia.org/10-5861-ijrsl-2019-3012>. <https://doi.org/10.5861/ijrsl.2019.3012>
19. 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. <http://www.academypublication.com/ojs/index.php/tpls/article/view/tpls0806665673>. <https://doi.org/10.17507/tpls.0806.17>
20. 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. <http://www.jallr.com/index.php/JALLR/article/view/729>
21. Jankowska, A. & Zielińska, U. (2015). Designing a self-assessment instrument for developing the speaking skill at the advanced level. In M. Pawlak and E. Waniek Klimczak (Eds.) *Issues in teaching, learning and testing speaking in a second language*. Heidelberg: Springer. https://doi.org/10.1007/978-3-642-38339-7_16
22. Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144. <https://ila.onlinelibrary.wiley.com/doi/abs/10.1002/jaal.525>. <https://doi.org/10.1016/j.edurev.2007.05.002>
23. Keshmirshakan, M. H., Namaziandost, E., & Pournorouz, M. (2019). The Impacts of Creative Writing Activities on Iranian EFL Learners' Progress in Writing, Writing Dispositions: Focus on Attitude to English Course. *EPH - International Journal of Educational Research*, 3(9), 12-22. <https://ephjournal.com/index.php/er/article/view/1552/911>
24. 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. <http://www.ccsenet.org/journal/index.php/elt/article/view/5253>. <https://doi.org/10.5539/elt.v3n1p184>
25. Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing*, 19 (1), 3-31. <https://journals.sagepub.com/doi/abs/10.1191/0265532202lt2180a?journalCode=ltja>. <https://doi.org/10.1191/0265532202lt2180a>
26. Llach, M. (2011). *Lexical errors and accuracy in foreign language writing*. New York: Multilingual Matters. <https://doi.org/10.21832/9781847694188>
27. Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511733017>
28. Metruk, R. (2017). Pronunciation of English dental fricatives by Slovak University EFL Students. *International Journal of English Linguistics*, 7 (3), 11-16. <http://www.ccsenet.org/journal/index.php/ijel/article/view/64987>. <https://doi.org/10.5539/ijel.v7n3p11>
29. 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. <http://www.academypublication.com/ojs/index.php/jltr/article/view/jltr090612701276>. <https://doi.org/10.17507/jltr.0906.16>
30. 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. <https://eric.ed.gov/?id=EJ1135930>
31. 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. <http://www.jallr.com/index.php/JALLR/article/view/1023>
32. 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. <http://www.jallr.com/index.php/JALLR/article/view/1031>

33. 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. <http://www.academypublication.com/ojs/index.php/jltr/article/view/jltr09018089>
34. 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>.
35. Namaziandost, E., Abedi, P., & Nasri, M. (2019). The Role of Gender in the Accuracy and Fluency of Iranian Upper-intermediate EFL Learners' L2 Oral Productions. *Journal of Applied Linguistics and Language Research*, 6(3), 110-123. <http://www.jallr.com/index.php/JALLR/article/view/1018>
36. 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. <http://jurnal.untirta.ac.id/index.php/JELS/article/view/5431>
37. Namaziandost, E., Nasri, M., & Rahimi Esfahani, F. (2019a). Pedagogical Efficacy of Experience-Based Learning (EBL) Strategies for Improving the Speaking Fluency of Upper-intermediate Male and Female Iranian EFL Students. *International Journal of Research in English Education*, 4(2), 29-41. <http://ijreeonline.com/article-1-160-en.html>. <https://doi.org/10.29252/ijree.4.2.29>
38. Namaziandost, E., Nasri, M., & Rahimi Esfahani, F. (2019b). Texts with Various Levels of Hardness, Reading Comprehension and Reading Motivation: I+1 Versus I-1. *ELT Forum: Journal of English Language Teaching*, 8(1), 60-77. <https://journal.unnes.ac.id/sju/index.php/elt/article/view/30992>
39. Namaziandost, E., Rahimi Esfahani, F., Nasri, M., & Mirshekaran, R. (2018). The Effect of Gallery Walk Technique on Pre-intermediate EFL Learners' Speaking Skill. *Language Teaching Research Quarterly*, 8, 1–15. https://www.researchgate.net/publication/328449149_The_Effect_of_Gallery_Walk_Technique_on_Pre-intermediate_EFL_Learners%27_Speaking_Skill. <https://doi.org/10.17507/tpls.0812.16>
40. 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. <https://journals.aiac.org.au/index.php/IJALEL/article/view/2803>. <https://doi.org/10.7575/aiac.ijalel.v.6n.1p.311>
41. 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. <https://journals.aiac.org.au/index.php/IJALEL/article/view/4289>. <https://doi.org/10.7575/aiac.ijalel.v.7n.3p.201>
42. 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. <https://ideas.repec.org/a/asi/ijells/2019p99-109.html>. <https://doi.org/10.18488/journal.23.2019.83.99.109>
43. 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.
44. Pan, M. (2016). *Nonverbal delivery in speaking assessment. From an argument to a rating scale formulation and validation*. Singapore: Springer. <https://doi.org/10.1007/978-981-10-0170-3>
45. Pokrivčáková, S. (2010). *Modern teacher of English*. Nitra: ASPA.
46. 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. <https://jurnal.ar-raniry.ac.id/index.php/equality/article/view/1448/0>
47. 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.
48. Saritha, K. (2016). Rubric for English language teaching research. *Research Journal of English Language and Literature*, 4 (2), 725-731. <https://scholar.google.com/citations?user=yiMY04wAAAAJ&hl=en>
49. 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), 18-31. <https://language-testing-asia.springeropen.com/articles/10.1186/2229-0443-4-5>. <https://doi.org/10.1186/2229-0443-4-5>
50. Shakibaei, G., Shahamat, F., & Namaziandost, E. (2019). The effect of using authentic texts on Iranian EFL learners' incidental vocabulary learning: The case of English newspaper. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 2(5), 422-432. <http://www.ijllt.org/the-effect-of-using-authentic-texts-on-iranian-efl-learners-incidental-vocabulary-learning-the-case-of-english-newspaper/>
51. 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.
52. Thornbury, S. (2005). *How to teach speaking*. Harlow: Pearson Education Limited.

53. 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>
54. 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>
55. Ur, P. (2012). *A course in English language teaching*. Cambridge: Cambridge University Press.
56. 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. <http://www.ijllt.org/linguistics-sla-and-lexicon-as-the-unit-of-language/>

Appendix A

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

Appendix B

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