The study of teacher Raters'on speaking perfomance assessment at senior high school in Kendari

Yeti[1], Aderlaepe[2], Sitti Agustina[3]
yetieti418@gmail.com
[1,2,3]*Halu Oleo University, Indonesia*

ABSTRACT

The objective of this study was to find out the study of teacher raters' on speaking performance assessment on students at senior high school in Kendari. This study used a quantitative design. The instrument used was students' public speaking in English Expo 2018. The population of the study was the English teachers in senior high school in Kendari, while the sample was 10 teachers which consisted of 5 males and 5 females' senior high school teachers of English from different schools taken by using simple random sampling. The data was collected by showing the video of the students' speaking performance from English Expo 2018 to the English teachers (raters) who were both male and female raters. The data was analyzed quantitatively by using SPSS version 23 to test the data normality and hypothesis. The hypothesis was tested using paired sample t-test to decide if the hypothesis was accepted or rejected. The findings showed that the outcome of data analysis showed no significant difference between the two genders. It could be seen the result of sig 0.102, which was greater than the alpha value of 0.05. This indicated that the H0 was accepted and H1 wa rejected. Therefore, it could be concluded that there was no significant difference in the average value given by male and female raters at senior high school in Kendari.

Keywords:
Raters' Gender, Speaking Achievement, Gender.

1. Introduction

Assessment is an important part of the language learning process. According to (Brown, 2004), assessment is a continuous activity that spans a wide range of topics. Assessing language learning is considered crucial, especially to evaluate whether the learning process has been conducted properly or if the students have achieved the target of language learning. Speaking assessment has also received little attention, despite speaking being a part of language skills. At school, teachers generally evaluate students' speaking ability through daily assignments. Moreover, students who actively address or answer questions during class are the ones who will receive high scores in speaking. The purpose of developing speaking skills in language learning is to provide students with the necessary tools to express their thoughts and opinions confidently. It aims to cultivate a mindset where students feel comfortable speaking up and actively participating in classroom discussions and speaking practice activities (Rahmasari et al., 2022). On the other hand, those who participate less in class will be considered to have a low score in speaking.

However, students' ability in speaking cannot be determined solely by observing them in the classroom. Speaking is the most essential things than the other skills in language learning, because it proves students' ability in producing language learning (Manu & Kamlasi, 2022). The situation that arises in speaking classes is that students know the answer but are reluctant to respond. Brown & Yule in Novitasari et al. (2022), The primary objective of advancing speaking skills to a higher level is to equip students with the ability to effectively communicate in the target language. This includes being able to express oneself, engage in basic interactive skills such as greetings and apologies, and convey needs, requests, and gather information. Additionally, many students lack confidence in speaking simply because they have seen their friends who are more fluent. Teachers should pay more attention and consideration before assigning scores to students. Therefore, assessment is necessary for students' learning process.

Speaking assessment itself is always in subjective form since it involves teachers grading individuals' speaking abilities. According to Fulcher et al. (2011), there are so many factors affecting our impression about how people are able to speak a language so the speaking assessment can be challenging, and students are demanded to produce answers completely in integrating language skills. In standardized tests, teachers use a speaking rubric as the basis for assessing students. Raters have the final say in determining whether test-takers/students are fluent enough in speaking, which ultimately determines their pass or fail status. Furthermore, raters' bias can stem from their background, their relationship with the test-takers, and even their gender (Johnson & Lim, 2009). Despite the crucial role they play in speaking assessment, it is often observed that there is rater bias in the assessment process. Raters may consistently assign lower or higher scores to students whose actual ability does not align with the level of speaking performance represented by the score, thereby invalidating and biasing the measurement. Consequently, students are considered to have mastered speaking in English when they can speak the language, rather than being assessed on their ability to comprehend the language.

Undoubtedly, as humans, subjectivity in evaluation makes us susceptible to inconsistent, inaccurate, and biased judgment patterns, which are known as rater effects in educational assessment. According to Wolfe & McVay in Eckes (2012), raters influence the ranking pattern of measurements, leading to errors in score validation. For instance, two raters may assign the same or different scores to the same speaking performance based on the same standards in a speaking test, but their expectations and interpretations of the performance may differ. Moreover, the opinions of raters and teachers can vary depending on their gender. For example, differences in conversation patterns between males and females can impact their assessments. Therefore, the way both genders speak may differ.

The results of observations conducted by the researcher in several high schools in Kendari revealed a phenomenon where students assessed by female raters mostly scored higher than those assessed by male raters. According to Aryadoust (2016), differences in interviewing styles between male and female raters can be considered as gender effects on assessment, influencing the way they assess language tests. Females, who prioritize grammar or accuracy, may place more emphasis on these aspects during assessment. They may have higher standards for assessing grammar.

On the other hand, males may focus more on paralinguistic competence, such as enthusiasm displayed through gestures, and may not view grammar as the most crucial part. Therefore, it is reasonable to assume that the gender of the raters in scoring poses a challenge in identifying the impact of rater gender on oral proficiency assessment.

2. Methods

The design of this study was a quantitative research design. Creswell (2014) stated that quantitative research design was a kind of research used to a t-test the theories where the number of data types on instruments could be analyzed by using statistical procedures. Population is defined as consisting of all members of a well-defined group (Lomax, 2001). Thus, the population of this study was the English teachers in senior high school in Kendari. Meanwhile, sample is defined as consisting of a subset of the population (Sugiyono, 2014). The sample in this study was determined by using a simple random sampling method. The sample of this study was English teachers of senior high school from different schools, which consisted of 5 male teachers and 5 female teachers. Therefore, the sample had to fulfill some criteria, including the teachers' teaching experience was 5 years (min) and had certifications.

The students' public speaking in English Expo 2018 was used as the instrument. The purpose of the instrument was to measure the ability of students toward their grammar, pronunciation, compensation techniques, the content of the utterance, language appropriateness, and overall intelligibility. This study also used a speaking test, speech samples (it was acquired from public speaking video competition from English Expo). In this phase, the video of speaking performance was shown to the English teachers (raters) both male and female raters. Students' speaking was assessed by the raters using a speaking rubric developed by Rooney (2003), which consisted of seven aspects of assessment in speaking, including gestures, the use of eye contact, the use of language, etc. Students' scores were categorized into four-level; excellent, good, satisfactory, and needs improvement.

After collecting the data, the data were analyzed quantitatively by using SPSS 23 version. The researcher applied the descriptive statistic for analyzing mean, standard deviation, maximum score, and minimum score. Besides, the inferential statistic used to test the hypothesis of this study (H_1). The researcher used paired sample t-test to test the hypothesis of this study, which was calculated through SPPS 23 version.

3. Results

In order to examine the raters' assessment of male raters and female raters, a one sample Kolmogorov-Smirnov test is used to see if the data are normally distributed which resulte of the male raters' Asymp. Sig. (2-tailed) is 0.724 and the female raters' Asymp. Sig. (2-tailed) is 0.995 that show the data are distributed normally. Then, in this section, the difference in assessment between male raters and female raters by the average score of the first student can be seen below:
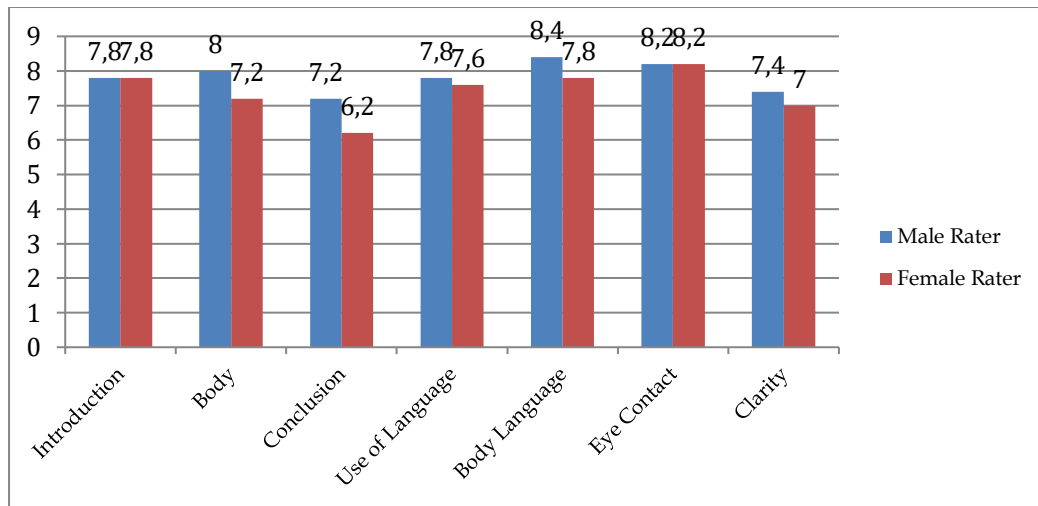
Graph 1. The comparison of average score between male and female raters on the 1st student

Based on the graphic above, it can be seen the difference of scoring between male and female raters for first students (female students). In the first criteria, namely the introduction, it can be seen there is no significant difference of score given by male and female raters, both of teachers give score of 7.8. In body criteria, male raters give a higher score of 8. Meanwhile, female teacher give lower rating with a score of 6.2. For the conclusion criteria, male raters give a higher rating with a score of 7.2. Meanwhile, female raters give a lower rating with a score of 6.2. Furthermore, for the use of language criteria, it can be seen that there is no significant difference between the scores given by male raters and female raters. Whereas, male raters give a score of 7.8 and female raters give a score of 7.6. In the body language criteria there are differences in the scores given by male raters and female raters. Male raters give a higher score than female raters with a score of 8.4 points, while female raters give a lower score with 7.8 points. Furthermore, for the eye contact criteria, both male and female raters give a score of 8.2. Likewise with the last criteria, that is clarity, there is no significant difference in the value given by male raters and female raters, namely with a value of 7.

Therefore, male raters give higher scores than female raters. All male raters gave an average score of 8 while the female raters gave an average score of 7.
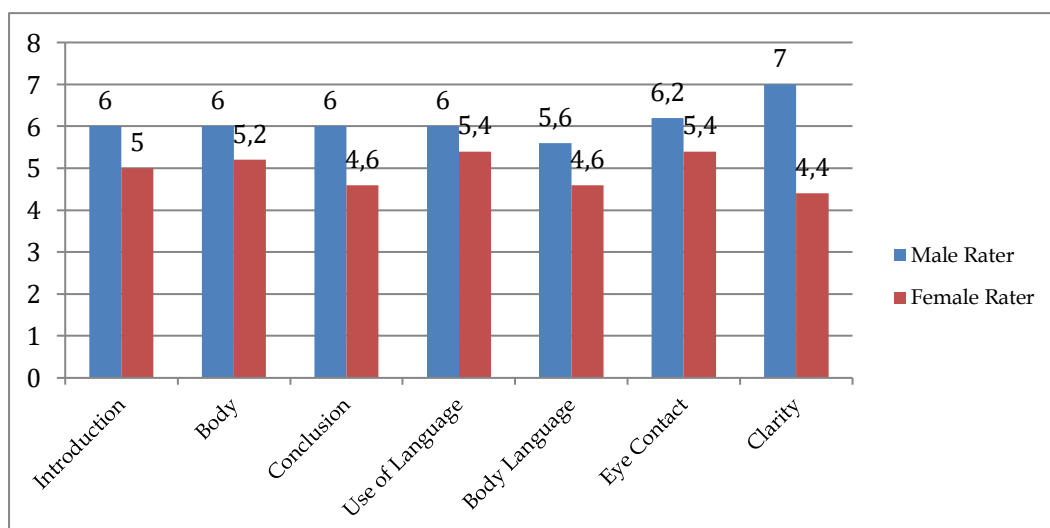
Graph 2. The comparison of average score between male and female raters on the 2nd student

The graph above shows a comparison of the scoring given by male raters and female raters to second student (male student). For the first point of introduction, male raters give an assessment of 7.8, while female raters gave a lower rating with a value of 5. This shows that there is difference between the scoring given by male raters and female raters. In the second criterion, namely the body also shows a little difference of the score given by male raters and female raters, for these criteria male raters give higher rating, than female raters with a score of 6. Meanwhile, female raters give a score of 5.2. The differences are also shown in the third criteria, namely the conclusion. Male raters give a higher rating of 6 while female raters give a lower rating of 4.6. For the use of language criteria, male raters give score of 6 and female raters give a score of 5.4. For body language criteria, there are differences in the score given by male and female raters. Male raters give a higher score than female raters with a score of 5.6 while female raters give a lower score with 4.6 points. Furthermore, for eye contact criteria, the scoring given by male raters is higher than female raters with a score of 6.2 while female raters give score of 5.4 points. Likewise with the last criterion, namely clarity, there are differences in the values given by male and female raters. The male raters give a higher score than the female raters with a score of 7 points, while the female raters give a lower score with 4.4 points.

Therefore, the graphs show that there is a slight difference in the scores given by male and female raters to female students. That is male rater give higher grades to students than female rater although the difference is not very significant.

4. Discussion

Based on the result of t-test for equality of means between male and female raters, the outcome of the table indicates that there was no significant difference between male and female raters in their scorings of students' speech performance in using each category of the rating scale. The result of the study also shows that there is no significant difference in the scores given by male raters or female raters. Both male raters and female raters give high ratings to second student. This finding is different from the statement from O'Sullivan (2000) that in his study investigated the impact of raters' gender differences in the assessment. The results showed that test takers scored higher when rated by female raters regardless of the test taker's gender. While the others revealed Caban (2003) showed different result where the test takers scored higher when male raters' assessed them. Therefore, in the present study, there is no big difference in the raters' gender when raters assessed test takers.

On the other hand, the findings of this research are both in line and in contradiction with previous research studies which found both no significant and a significant between female rater and male rater. The previous research conducted by Bijani & Khabiri (2017), they are also researching the relationship between the impact of raters' gender on the awarded scores to males and female test takers. The results of their research state demonstrated there is not any significant role in their performance differences when they were rated by the raters of the same or opposite gender. Moreover, no significant difference was observed regarding male and female raters' biases towards the rating scale categories. Related to research conducted by

Motallebzadeh & Nematizadeh (2011) showed that through the t-test it was found that females were better in oral performance than males. However, the difference was not very significant. It can be concluded that male and female students have differences, but the differences are not too significant. Therefore, the results showed that there is no difference in the average raters' male and female raters' in giving grades to students. Related to research conducted by O'Sullivan's (2000) research related Exploring gender and oral proficiency interview performance. The results of the study showed indicated systematic gender deference, while analysis of the responses of the learners suggests a tendency to produce more grammatically accurate language with their female interviewers. Therefore, it can be concluded that there are significant differences between male and female raters although only slightly different.

Therefore, the findings of this study suggested that gender, either on account of the raters or the test takers, does not have any significant impact, on the one hand, on the performance ability of the test takers. In other words, test takers' gender differences do not affect the quality and quantity of their oral performance. Besides, raters' scoring is not affected by the same or opposite gender since the data did not show any bias caused by gender differences. The outcomes of this research provide further evidence on the similarity of male and female rater performances in assessing. It can be seen from 6 out of 7 aspects which are introduction, concluding, and the use of language, body language, eye contact, and clarity.

In the end, the result of this research showed no significant difference when the English teachers acting as raters in assessing students' speech performance on English Expo 2018. Even if there are differences in assessing, it is due to the students' speech performance that was showed by the researcher. It can be said that the English teachers that acted as raters are giving assessment fairly even if they have different gender.

5. Conclusion

Based on the data analysis above, this study can be concluded that male raters give a higher scores than female raters. So the result of the study indicates that there was no significant difference between male and female raters in their scorings of students' speech performance in using each category of the rating scale. The result of one sample t test that shows the result of sig 0.102 which is greater than the alpha value of 0.05, makes H_0 accepted that there is no significant difference for the average value given by male and female raters. The result of the study also shows that there is no significant difference in the scores given by male raters or female raters. The findings of this study suggested that gender, either on account of the raters or the test takers, does not have any significant impact, on the one hand, on the performance ability of the test takers. In other words, test takers' gender differences do not affect the quality and quantity of their speaking performance. Besides, raters' scoring is not affected by the same or opposite gender since the data did not show any bias caused by gender differences. The outcomes of this research provide further evidence on the sameness of male and female rater performances in rating. Therefore, there will be no excuse on the side of decision makers to exclude the raters of either gender from rating test takers' oral performances. The findings of this study are both in line and in

contradiction with previous research studies which found both no significant and a significant between female rater and male rater.

References

Aryadoust, V. (2016). Gender and Academic Major Bias in Peer Assessment of Oral Presentations. *Language Assessment Quarterly An International Journal*, *3*(1), 1–24. https://doi.org/http://dx.doi.org/10.1080/15434303.2015.1133626

Bijani, H., & Khabiri, M. (2017). The Impact of Raters' and Test Takers' Gender on Oral Proficiency Assessment: A Case of Multifaceted Rasch Analysis. *Journal of Teaching Language Skills (JTLS)*, *36*(3), 1–32. https://doi.org/http://dx.doi.org/10.22099/jtls.2017.24841.2223

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education, Inc.

Caban. (2003). *Rater group bias in the speaking assessment of four L1 Japanese ESL students* [University of Hawaii at Manoa]. http://hdl.handle.net/10125/40655

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.)*. Sage.

Eckes, T. (2012). Operational Rater Types in Writing Assessment: Linking Rater Cognition to Rater Behavior. *Language Assessment Quarterly*, *9*(3), 270–292. https://doi.org/https://doi.org/10.1080/15434303.2011.649381

Fulcher, G., Davidson, F., & Kamp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Journal of . Language Testing*, *28*(1), 5–29.

Johnson, J. S., & Lim, G. S. (2009). The influence of rater Language background on writing performance assessment. *Journal of Language Teaching*, *26*(4), 485–505. https://doi.org/http://dx.doi.org/10.1177/0265532209340186

Lomax, R. G. (2001). *An Introduction to Statistical Concepts for Education and Behavioral Sciences*. Lawrence Erlbaum Associates.

Manu, D., & Kamlasi, I. (2022). A Study on the Ability in Using Preposition of Place in Speaking Performance. *Journal of Teaching of English*, *7*(1), 81–92.

Motallebzadeh, K., & Nematizadeh, S. (2011). Does Gender Play a Role in the Assessment of Oral Proficiency? *English Language Teaching*, *4*(4), 165–172. https://doi.org/http://dx.doi.org/10.5539/elt.v4n4p165

Novitasari, Nurnia, & Kamaluddin. (2022). Students Emotional Quotient and Their Speaking Skills; a case of SMA Negeri 4 Baubau. *Journal of Teachingof English*, *7*(3), 83–88. http://jte.uho.ac.id/index.php/journal/article/view/22/16

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System*, *28*(3), 373–386. https://doi.org/https://doi.org/10.1016/S0346-251X(00)00018-X

Rahmasari, Kamaluddin, & Agustina, S. (2022). Teaching Speaking Skill By Using Probing Prompting Learning Strategy At SMPN1 Lawa. *Journal of Teachingof English*, *7*(3), 51–61. http://jte.uho.ac.id/index.php/journal/article/view/131

Rooney, J. E. (2003). Blending learning opportunities to enhance educational programming and meetings. *Association Management*, *55*(5), 26–32.

Sugiyono. (2014). *Metode Penelitian Kuantitatif, Kualitatid dan R&*. Alfabeta, CV.