An Investigation into English-Majored Students' Self-assessment of Their Speaking Performance

Khau Hoang Anh
Tra Vinh University, Vietnam, *anhkhau@tvu.edu.vn*

Mai Hung Dong
Tra Vinh University, *maihungdong99gmail.com*

Nguyen Huynh Trang
Dr., corresponding author, University of economics Ho Chi Minh, Vietnam, *trangnh@ueh.edu.vn*

Self-assessment with corrective training has been long seen crucial in helping learners improve their own work. This study investigates a group of 28 English-majored students' self-assessment of their Speaking performance at the end of a Speaking course at a university. First, the study explores their knowledge of the speaking test and then it provides them with 11 meetings of training on and practising using a speaking rubric which centres on the four grading criteria: Fluency and coherence, Pronunciation, Grammatical range, and Lexical resources. The participants' perception about their speaking performance after they took the speaking test is also explored. Finally, two examiners are employed to rate these students' speaking performance. The results revealed that the participants were found to be knowledgeable about the requirements in doing the test. Then, for their self-assessment on their test performance, they used the trained rubric to help them rate their work quite well, which highly correlated with the result of the teachers' rating which was also correlated with the average score of the two raters. The students' perceptions about their ability to perform these four Speaking grading criteria after taking the test also unveiled a strong connection to their overall self-assessment result.

Keywords: self-assessment, English-majored students, speaking performance, rubric, grading criteria, perceptions

INTRODUCTION

Self-assessment can be seen as a useful tool for learners to reflect themselves about doing a task. It needs to be done regularly to ensure constant progress and propose corrective measures for improvement if necessary. Self- assessment has been proved to be effective in many cases. In this context, it should direct us into a case of learning a foreign language. Butler and Lee (2010) found that 6th grade Korean students were able

to improve their capability to self-assess their own performance constantly in their English class. Huang (2016) also discovered that after training students how to self-assess and give feedback in a foreign language speaking test, they were able to provide diversified and satisfactory feedback. And many other researchers like Barney, et al. (2012); Wang (2016); Seifert and Feliks (2018) also advocated deployment of self-assessment in class.

Although many studies have supported the use of self-assessment to enhance someone's work, some concerns are mentioned. Butler (2016) said self-assessment can cause subjectivity and inadequate judgment about someone's task. These concerns can be true in many circumstances and therefore these have made us curious about the employment of self-assessment and in this study we would like to investigate if students in our class are able to assess their own speaking objectively and accurately after receiving the training of self-assessment.

More specifically, in this research, the researchers are trying to restrict themselves to exploring English-majored students' self-assessment of their speaking performance after finishing their speaking course. It is conducted because such an investigation is vital to having a look-back on their progress and the instructor's teaching method as well. No much similar research has been conducted to see if the test-takers really know what they are going to be tested on and to see how they self-assess their own speaking performance based on a set of criteria. Usually, test-takers are tested and the test-providers do not have adequate information about each individual. In order to facilitate this aim, two objectives are set. One is that the participants (test-takers) will be asked questions centering on the test requirements. This should be seen as fair assessment as this must be clearly presented at the beginning so test-takers do not feel confused about the test format. The second objective is to see how they perceive their capability in completing each grading criterion in the speaking test. The third one is to discover their own capability to self-assess their own speaking work based on a set of criteria. Finally, the study attempts to test the correlation between their self-assessment result and that of the examiners'. To have this study go in direction, the researchers would like to propose the four following research questions.

1. Are students knowledgeable about the speaking test they take at the end of their course?
2. How do they perceive their self-assessment of each speaking grading criterion?
3. How do they assess their own speaking performance on this test?
4. Is their self-assessment result correlated with that of the examiners?

Literature Review

Theoretical Framework

In any study, a theoretical framework should be clearly presented to help guide the implementation of a work, so the following framework should be employed to help orient this study by defining speaking, the format of the test, fluency, pronunciation, vocabulary, the rubric and perceptions. In Vietnam, English is treated as a foreign language and the means of instruction in this particular course is English.

First, what is speaking? Speaking is a helpful interaction skill in language performance, which requires more pronunciation skills. Furthermore, it can be implemented in three speaking interaction types: It can be face-to-face conversations or telephone conversations; Speaking situations are directly interactive for instance, making a speech to a live audience and speaking situations may be totally non-interactive, such as recording a speech for a radio broadcast (Reyad & Alhaj 2020). Speaking is a specific spoken discourse that is a great tool to communicate and engaged in for social purposes. It consists of three areas of knowledge: the mechanical elements of language (pronunciation, grammar, and vocabulary), which allows the speaker to use the right words in the correct sequence and appropriate pronunciation, a conversation's functions, such as a transaction and interaction. They also added speaking involves the socio-cultural norms like turn-taking, rate of speech, length of pauses between speakers, relative roles of participants (Burnkart, 1998 as cited in Hendra, 2012).

Second, as the study also uses the following test format, so the participants are supposed to know well about it. The study adapted the test format by Guy and Vanessa (2012) consisting of three parts as follows:

> Part 1: The examiner asks a number of questions about familiar topics such as the candidate's studies/work, hobbies, interests, etc. Part 2: After a minute's preparation, the candidate speaks for two minutes on a familiar topic provided by the examiner. Part 3: The examiner and the candidate discuss some general questions based on the theme of the Part 2 topic. Candidates are expected to be able to respond to questions on familiar and unfamiliar topics and to speak at length. Candidates are assessed on a nine-band scale for fluency, vocabulary, grammar and pronunciation. (p.7)

The participants will be trained on how to self-assess their speaking performance, so this study requires clear descriptions of each component in a speaking task.

Fluency

"Fluency is an impression on the listener's part that the psycho-linguistic processes of speech planning and speech production are functioning easily and efficiently" (Lennon 1990, as cited in NIVJA, 2011, p.2). Skehan (1998), considered fluency to be one of the most factors in the aspects of second language performance, namely speaking, reading, pronunciation, grammatical processing, and promote fluency (as cited in Shumei, 2009).

Pronunciation

"Pronunciation is one the expressions that a person shows his/her self-image to seek to change someone's pronunciation - whether of the first or second language - is to tamper with their self-image, and is thus unethical - morally wrong" (Porter & Garvin, 1989, as cited in Christiane & Barbara, 1994, p.7).

Grammar

"Grammar is the study of all the contrasts of meaning that it is possible to make within sentences" (David, 2006, p.32). According to Verghese (1989), grammar is made of

certain rules which control the organization of languages and structures. It gives us practice in using language (as cited in Ali & Farhad, 2016).

Vocabulary

"Vocabulary is the knowledge of words' meaning; words come in at least two forms such as oral and print. Knowledge of words also comes in at least two forms, receptive – that we can understand or recognize, and productive - the vocabulary we use when we write or speak." (Elfried et al. 2005, as cited in Mona, 2013, p.88). Hamad (2013) posited that vocabulary plays a crucial role in speaking performance. It is one of the factors that has affected speaking performance greatly and can lower a candidates' speaking performance (as cited in Raja et al., 2018).

Rubric

According to Andrade and Du (2005), rubric is a "document that articulates the expectations for an assignment by listing the criteria, or what counts and describing levels of quality from excellent to poor" (p.1).

To support the theory of rubric, Glenn (2005) said

> The assessment rubric is a descriptive schema, used as a benchmark in analysing the product as well as the business process and the achievement of the student. Thus, the rubric can reveal "how students learn, what students should learn, how student learning is assessed, and how assessments are generated (as cited in Suryanti & Nurhuda, 2021, p.670).

Perceptions

According to Ou (2017), perception is "the process of attaining awareness or understanding of sensory information. The word "perception" comes from the Latin words, meaning 'receiving, collecting the action of taking possession, and apprehension with the mind or senses". Colman (2006) said that "perception is defined as an act, process or product of perceiving to become aware or gain knowledge of something through the senses" (as cited in Pham, 2020, p.197). As Nikian et al. put it, "perception is regarded as an individual interpretation of something which is affected by past experiences, current understanding, present situation, and information" (as cited in Budasi et al. 2020, p.512).

With these concrete theories, the study affirms to be conducted in a right direction.

Related studies

This current study aims at exploring whether or not a group of English-majored students is able to assess their speaking performance and what they think about their own performance based on the trained rubric in class, so the study attempts to find what previous studies found about the use of rubric and the participants' perceptions about using it in the classroom.

Su (2021) recruited a group of 32 third-year students in the English Department to offer training on interpreting in a period of three weeks. The study used a skill-based

assessment form (English-Chinese Sight Interpreting). During self-assessment, they were asked to look for descriptors in the can-do-statements for what they had mastered. They then gave some remarks on their performance based on these chosen descriptors. They needed to use examples from their interpretations to justify their previously provided remarks. It revealed that after the training of using this rubric, the students acknowledged better the interpreting features thanks to their familiarizing themselves with descriptors and appropriate comments with evident examples to illustrate the comments. They better knew the skill of cross-language transference and source language reference, discourse devices, and some other related features. So, clearly stated descriptors allow the students to do the job better as these help guide them all way through their work completion.

Thorben, et al. (2020) made an investigation to see if any difference in judgment between experienced teachers and student teachers. Both groups assessed four essays written by four students, using a holistic and a separate analytical scale. The holistic result was compared with that of a machine rating score while the result of analytic scales was compared with that of expert raters. It revealed that experienced teachers showed stricter ratings. They looked for more detailed errors while student teachers seemed less strict or they might not experience much in error detection. Although so, the results of both groups were still more negative than machine rating and expert rating. This can be true in the real-life classroom where the same teacher can rate his students' same work differently at different times in a day or in a course. It is even worse when works are rated by different teachers with distinctive qualities or experience. In this current study, two raters are used to rate the participants' speaking performance as it hopes to bring a reliable result.

Farooq et al. (2020) shared a single essay written rubric for both the teacher and students with criteria like organization, language, and vocabulary and they also used a questionnaire attempting to elicit the students' attitudes for self-assessment. The authors of this study used the same cycle to assess an essay in class flowing like this: Writing an essay- students' self-assessment - backwash effect - teacher assessment in three weeks. The students were trained on applying this cycle. The result revealed that the students could evaluate their essay with instructions and guidance. They tended to use the rubric with little help afterward. They gave more marks for their essays in the first time assessment than the second time. When looking into inter-rater reliability between the students' assessment and the teacher's assessment, for the first time, they showed a huge statistical difference. In the second time, both assessments were still statistically different, but the gap was smaller due to the students' familiarity with the rubric. In terms of teachers' ratings in the first time and the second time, it was found that the results were no statistical difference in assessment. So, when the students are facilitated with times to access to the rubric, they are able to use it more effectively afterward.

Kim (2019) offered training of using a rubric-referenced self-assessment to a group of nineteen 11th grade Korean students to explore if this worked in her class and how they thought about the rubric. The students were asked to draft their essay multiple times based on the provided rubric and the progress was seen in the consecutive drafts. With

this strong impact on their essay quality, the students found it interesting in using the rubric and showed their positive attitudes towards using this as a learning strategy to improve their essays. So, this group of high school students is also able to optimize the use of rubric to facilitate their learning process.

To explore the effects of peer assessment on EFL students' English - Speaking performance and learning perceptions, Chien et al. (2019) employed 69 Taiwanese high school students from two classes of a high school for the study. 33 students experimented. None of them had any previous experience of using peer assessment in English-speaking courses. The assessment was implemented on three grounds. The first ground showed no statistically significant correlation between peer scores and teacher scores. The second ground with ($r = 0.69$, $p = 0.00 < 0.01$, $r2 = 0.48$) showed a statistically high correlation and in the third ground the scores between the two groups were statistically moderately correlated. It also disclosed that they perceived peer assessment as less stressed and helped improve learning. This study supports the use of rubric in helping enhancing a group of high school students and as can be seen when the students are used to the rubric, they tend to become expert users of it.

Seifert and Feliks (2018) made an investigation on self-assessment and peer assessment to help improve students' and teachers' assessment skills. The research sample included 300 students studying for Bachelor's or Master's degrees in education. They used a set of questions to assess the students' perceptions about self-assessment. It unveiled that the students regarded self-assessment as fair work with a mean score of 4.4 and they confessed that self-assessment helped contribute to the enhancement of their work ($M=3.7$). They found the participants were more responsible for their assessment, knew assessment complexity, and needed investment in it. They consented that teachers should make use of both holistic and analytical rubrics to help learners know and practice so that both students and teachers improve their learning and teaching. Making use of the rubric in this case helps the classroom teacher and students improve their teaching and learning.

Wang (2016) conducted a research on a group of 80 university students on their perceptions about using rubrics for self-assessment in EFL contexts. It revealed that these students showed positive perceptions on the deployment of rubrics in self-assessment in their class. They considered that rubrics help promote their self-regulation and self-reflection in their learning, which was in support of empowering their writing performance later on. So, this study continues to advocate the use of rubrics.

In short, many studies investigated students' use of rubric in helping guide them through work completion and most found that the participants were able to make use of it. What is more, the student participants tended to tolerate their own work at least at preliminary stages of practicing using the rubric. Another experience is that they were happy with the use of rubric and could become expert users of it later on. However, this current study hopes to see the result of the students' self-assessment on their own speaking performance and their perceptions about what they did in the speaking test after practicing using the rubric in a period of 11 weeks of instruction.

METHOD

Research design

To collect the data for the research questions, the study employed a group of third-year English-majored students in a speaking class and two examiners/raters to be the participants. Then the speaking test format used in the study was adapted from Guy and Vanessa (2012). The set of criteria for assessing the students' speaking performance was adapted from Wattie (n.d.) (Please refer to the Appendix). To have a fair assessment, the study also aims at testing the students' knowledge about the test requirement in addition to the rubric they practiced in class. More detailed information will be presented as follows.

Participants

The participants are a group of English-majored students who were following a Speaking course in Tra Vinh University. They were chosen as they had been trained on using the speaking rubric since the first class. In addition, they were trained on how to perform different speaking tasks sticking on the IELTS-format. They joined the study voluntarily as their participation brought them more experience in doing such tasks and they also had a better opportunity to practise using the rubric with familiar grading criteria. Their mother tongue is Vietnamese and they learn English as a foreign language. Two English teachers were recruited to assess these students' Speaking performance. One of them used to attend a one-month training course in grading speaking and writing in English in Hue University, Vietnam conforming to the requirement of the Vietnamese Standardized Test of English Proficiency.

Instruments

To collect the data for the first research question, the present study applied a questionnaire concluding 16 questions about their knowledge of the Speaking test. The content questions were developed based on the work written by Guy and Vanessa (2012) and then the questionnaire was delivered to the participants for their knowledge of the Speaking test, using a five-point Likert scale (1: Strongly disagree, 2: Disagree, 3: neutral, 4: Agree, 5: Strongly agree). The reliability of the responses obtains .804, which can be used for analysis as shown in Table 1 below.

Table 1
Reliability of the questionnaire

Cronbach's Alpha	N of Items
.804	16

Another instrument, also a questionnaire, sought students' perceptions about their self-assessment on the four grading criteria in their speaking performance, using the same scale as the first questionnaire and they were based on the work proposed by Wattie (n.d.). The third instruments utilized the self-assessment result provided by the participants themselves and the scoring results provided by the two examiners. The results of the two raters were correlated to see if any gap and then the scores of the examiners were averaged for correlating with that of the students. Two raters were used

because this is the school regulation, so instead of employing only one examiner as said in the theoretical background, an IELTS - like style, the study employed two as a Cambridge - like style. Instead of using a nine-band score, an IELTS - like style, the researchers used a 10-scale scoring system stipulated in the system of Vietnam education.

Procedure

The classroom teacher's instruction can play an integral part in ameliorating students' ability to monitor their learning. To continue this idea, the instructor also one of the researchers in this present study deployed a teaching method which we agreed upon on called "Monitor Your Work", which means using the rubric to improve students' Speaking performance. This was implemented with the following steps. In a period of 11 weeks (3 hours a week/meeting) of instruction using this method, the instructor trained his students on practising the content and the format of their Speaking test to be tested, which would take place at the end of their course. The instructor showed his students the format of the Speaking test in their course book and the speaking criteria used to rate their Speaking test (please relate to it in the Appendix). In each class meeting, the students were told to sit in groups or pairs to practise using this format and rating each other's work based on these criteria. Next, the instructor started to give them the topic to practise questioning and answering the questions in the given topic. They usually swapped their role for each turn. The students were told to give their teammate/s a band score and an analytical score on each criterion. Through observation in many class meetings, the students felt excited when they were told to stand up and give their teammates some remarks after group/pair practice. Then, the students who received the remarks felt not less excited and were told to think about their peers' remarks which could be right or wrong about them. In each class meeting, the instructor often invited a group of three to stand up (one playing the interlocutor, one the assessor, and the left playing the candidate) and he recorded this conversation and then would play it loud with the loudspeaker later on. The whole class listened to them carefully and referred to the rubric worksheet to see how this candidate worked. Some of the audiences were invited to give this candidate a score, too. Finally, the instructor directed the whole class to use the rubric while listening to this candidate's recording again, gave him or her a score plus remarks. The cycle progressed like this in every class. In the last class meeting, the participants were invited to join a survey about their knowledge of their upcoming Speaking test. Then, the instructor told them to bring the rubric worksheet with them on the test date because they needed to use it to rate their performance after their test completion. After that, they were asked to take a survey to give their perceptions about their performance in the test on the four speaking domains. Finally, the instructor and his partner collected all these worksheets for analysis afterward.

Data processing

The data were analyzed using the SPSS software, version 26 (Statistic Package for the Social Sciences). The researchers looked for internal reliability of the survey responses, run Descriptive Statistics Inter-rater correlation, and self-rating results. Then, they started looking for a comparison. The researchers averaged the inter-raters' scores

using a scale of 10 focusing on fluency and coherence (25% of weight = 2.5 marks), pronunciation (25% of weight = 2.5 marks), grammatical range, and accuracy (25% of weight = 2.5 marks), and lexical resources (25% of weight = 2.5 marks), respectively.

RESULTS

Table 2 below presents the descriptive analysis of the participants' knowledge of the speaking test. 16 questions were used asking them about the test format and other requirements for the test and this was done to answer the first research question.

Table 2
Students' knowledge of the speaking test

Item	Statements	N	Min	Max	Mean	Std. Dev
1	I know that there are three parts in the Speaking test.	28	2.00	5.00	4.3929	.78595
2	I know that the questions in the Speaking test don't ask a candidate about a specialized subject/field.	28	1.00	5.00	3.4643	.83808
3	I know that to do the Speaking test better, a candidate must master the grading criteria.	28	1.00	5.00	3.6786	1.12393
4	I know that Part 1 asks a candidate about general questions that are common in daily life.	28	2.00	5.00	4.2500	.84437
5	I know that in Part 1 a candidate needs to answer questions on three different topics.	28	2.00	5.00	3.6429	.98936
6	I know that Part 1 usually contains short questions which are seen as a warm-up.	28	3.00	5.00	4.0357	.74447
7	I know that Part 2 asks a candidate to describe a randomly selected topic which is a story-like genre.	28	2.00	5.00	3.9643	.74447
8	I know that Part 2 gives a candidate 1 minute to prepare the answer for the topic.	28	2.00	5.00	4.2857	.80999
9	I know that in Part 2 a candidate is allowed 1-2 minutes to talk about the topic.	28	3.00	5.00	4.4286	.69007
10	I know that the topics in Part 2 are about common topics, such as holidays, accommodation, and entertainment.	28	1.00	5.00	3.5000	1.10554
11	I know that Part 3 asks a candidate some more questions related to the selected topic.	28	2.00	5.00	4.1786	.90487
12	I know that Part 3 contains more challenging questions, using longer and more critically demanding questions.	28	2.00	5.00	4.0714	.94000
13	I know all the grading criteria for the Speaking test.	28	1.00	5.00	3.5000	.92296
14	I know that in general in order to be awarded a higher band score, a candidate must stick to the requirement of each component (fluency and coherence, pronunciation, vocabulary, and grammar).	28	2.00	5.00	4.0714	.94000
15	I know that in Part 2, a candidate's band score will be hurt if they speak less than the required minute.	28	2.00	5.00	3.8214	.90487
16	In both Part 1 and 3, a candidate's band score will be hurt if they cannot elaborate their answers.	28	3.00	5.00	4.0000	.72008
	Valid N (listwise)	28				

The mean values of this set of items range from 3.46 to 4.42, respectively and the standard deviations range from 0.72 to 1.12. As shown in the table, nine out of sixteen items obtained the mean score of over 4, meaning that they know the information of the Speaking test thoroughly except that Item 2 received the mean score below 3.5, meaning that a majority of the participants felt uncertain about this item of knowledge.

Table 3 below illustrates the participants' perceptions about how they did in the speaking test. More specifically, they gave their perceptions about how they did on each grading criterion after they took the speaking test and this was done to answer the second research question.

Table 3
The participants' perceptions about the speaking grading criteria after taking the test

Criteria	Statements	N	Min	Max	Mean	Std. Dev
Fluency & coherence	I think I did well in the Fluency and coherence domain.	28	2.00	4.00	3.0714	.81325
Pronunciation	I think I did well in the Pronunciation domain.	28	2.00	5.00	3.2500	.75154
Grammatical range & accuracy	I think I did well in the Grammatical range and accuracy domain.	28	1.00	4.00	2.9286	.81325
Lexical resources	I think I did well in the domain of the Lexical resources.	28	2.00	4.00	2.9643	.69293

As can be seen from Table 3, the mean scores of their perceptions about their ability to perform the Speaking test on the four Speaking domains after taking the test were seen. The domain of Pronunciation was obtained (M= 3.25, SD=0.75), which was the highest in all. Fluency and Coherence domain stood the second place with M = 3.0714, SD =.81325. The remaining domains appeared to be neutral on a five-point Likert scale.

In order to collect the data for the third and fourth research questions, first the mean scores for each criteria extracted from students' self-assessment are presented in Table 4 below. It should be reminded here. A 10-scale score is used for the overall test and each criterion is awarded 2.5 points.

Table 4
The students' self-assessment of the four grading speaking criteria

Criteria	N	Min	Max	Mean	Std. Deviation
Fluency & coherence	28	1.75	2.25	1.9693	.14247
Pronunciation	28	1.50	2.25	1.9566	.17078
Grammatical range & accuracy	28	1.75	2.25	1.9248	.15000
Lexical resources	28	1.50	2.25	1.9068	.17924

As can be seen in Table 4, the mean score of each of the grading domains weighs 25% in an overall score of a 10 - scale scoring system. This means the highest score for each of the domains receives 2.5 marks. As first glance, the Fluency and coherence domain obtains the highest mean score of 1.9693 with SD=.14247 and was closely followed by

Pronunciation (M=1.9566, SD= .17078). The two remaining criteria showed above-average performance.

Then, the scoring results provided by the two raters are illustrated in Table 5 in the form of correlation.

Table 5
Correlation of the two raters' scores

		Fluency & coherence R1	Pronunciation R1	Grammatical range & accuracy R1	Lexical resources R1
Fluency and coherence R2	Pearson Correlation	,959**	,963**	,950**	,957**
	Sig. (1-tailed)	.000	.000	.000	.000
	N	28	28	28	28
Pronunciation R2	Pearson Correlation	,956**	,969**	,963**	,956**
	Sig. (1-tailed)	.000	.000	.000	.000
	N	28	28	28	28
Grammatical range and accuracy R2	Pearson Correlation	,905**	,928**	,966**	,942**
	Sig. (1-tailed)	.000	.000	.000	.000
	N	28	28	28	28
Lexical resources R2	Pearson Correlation	,935**	,951**	,948**	,969**
	Sig. (1-tailed)	.000	.000	.000	.000
	N	28	28	28	28

Note. Correlation is significant at the 0.01 level (1-tailed).

Table 5 shows a correlation of nearly +0,9. Comparing all criteria, the given table illustrates that rater R1 and rater R2 have highly significant correlations (sig = ,000, r = 0,9 with 90%). As a rule in the university, the speaking test must be conducted and supervised by the two raters whose scores were then averaged for an overall score for each student. The raw score for each domain is 10. Each domain score is worth 25% of the overall score.

In order to see if the raters' scoring result is correlated with that of self-assessment, Paired samples statistics was applied and the result is shown in Table 6.

Table 6
Paired samples statistics for raters' scoring result and students' self-assessment

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Raters' scoring result	7.7643	28	1.51033	.28542
	Students' self-assessment	7.7750	28	.59169	.11182

This data was used to compare raters' scoring result (M=7.7643, SD=1.51033) with students' self-assessment (M=7.7750, SD=.59169). As can be seen in Table 6, the mean scores of the 2 groups were nearly the same.

Table 7
Paired samples correlation between the two groups' rating

		N	Correlation	Sig.
Pair 1	Raters' scoring result & Students' self-assessment	28	.022	.912

Raters' scoring result and students' self-assessment relate strongly together proved in the given table above with r = 0.22 and p =.912.

DISCUSSION

The first part of the discussion is about the first research question, which sought to determine students' knowledge of the Speaking test. This should be done to be fair. If the students had not been informed of this, they might have been confused. As can be seen in Table 2 above, the participants felt uncertain about three knowledge items of the Speaking test. They are Items 2, 10, and 13, which gained below the mean score of 3.5, but are still acceptable. Fortunately, the other thirteen items showed positive results, achieving the means scores of over 3.5, meaning they were ready for the test. This test of knowledge is crucial as Su (2021) found in her study. Test type and rubric familiarity play an integral part in upgrading students' learning outcomes.

The second part of the discussion is about the second research question, which directs us to further analyze their self-assessment of their Speaking test. First, the survey questions on their perceptions about their performance in the Speaking test on the four grading criteria as seen in Table 3. They were just not quite certain about their performance in pronunciation and Fluency and coherence, accounting for just above level 3 (M=3.0714 and M=3.2500, orderly) in a five-point-Likert scale. They showed even less confidence in using grammar and vocabulary. The mean scores of these two made up 2.9643 and 2.9286 in order. Overall, their perceptions of these four domains were quite negative. Many of them might not trust their ability to do the test.

Then, when giving it a closer look at what they self-assess their test on a 10-scale-score system, they awarded themselves just an above-average score. As seen in Table 4, the mean score for 28 students reached 7.7750 with its standard deviation of .59169. This can entice us into believing that their perceptions and their self-assessment (with the rubric for reference) match quite well.

When comparing these students' self-assessment with their examiners' ratings, the researchers found that the scores given between the two groups were not much statistically different. While the examiners/raters' mean score reached 7.7643, SD=1.51033, the students' self-assessment mean score obtained 7.7750, SD=.59169. As can be seen in Table 6, the mean scores of the 2 groups were nearly the same. However, the standard deviation of students' rating is higher than raters' rating, approximately 1.00 (Table 7). This result revealed the better correlation between students' self-rating

and teachers' rating compared to the finding of Mahwish et al. (2020). Students rated their work close to the rubric for the second time, but still, have some gaps with their teachers. What is more, the result from students' self-assessment and their perceptions about their performance was quite close and even close to the raters' rating. The students in this study seemed too strict to their performance while the student teachers in Thorben, et al. (2020) were tolerant of their grades. Surprisingly, it possibly depends on the purpose of the test or the weight of the test, students might treat their self-assessment differently. For example, if the test result is just used to test what they have achieved in order to amend what they have not achieved yet, students might not tolerate themselves, but if the test result is the important one. For instance, if it is used to compete for a reward or a scholarship, they may give a better score for their own work. The participants in this study did not use their self-assessment result for their accumulated score, so they can be predicted to give a fair self-assessment which is quite close to that of the examiners. The academic accumulated score is given by the examiners. To further discuss, although the present study did not investigate peer assessment, it can, in fact, indirectly influence this result. As seen in the Literature, the cycle of using the rubric involved peer's remarks, so this also matched the way Chien et al. (2019) did in their study when peer scores were moderately correlated with teacher scores. In addition, Kim (2019) found students felt interested in the use of a rubric-referenced self-assessment which was also seen in this study. In class, many students laughed with one another enthusiastically when they were listening to their peers' and teacher' comments.

In short, with these data, it can come to conclude that a careful lesson plan for instruction is extremely vital to the success of students. These students familiarize themselves with the format of the Speaking test and common topics to be tested in the test. They had also been trained to use the rubric in class and were given opportunities to sufficiently practise using the rubric to rate each other's speaking tasks. More importantly, the instructor had given them remarks on their tasks based on analytical criteria and given them a score. The time for students to practise using the rubric should be also brought into consideration. In this study, the students had many opportunities to repeat the reference of the rubric to give comments on each other's work and on their own work. This job can help imprint in their mind with such a rubric and remarks. This is in line with Barney, et al. (2012), who investigated the usefulness of rubric-based self-assessment as long as the student takes more time for their rubric investigation, kind of sufficient rubric exposure.

CONCLUSION

The present study purposed to investigate student's knowledge of the Speaking test and their rating ability in their Speaking test. The findings indicated that the majority of participants have mastered the knowledge of the Speaking test. In addition, to see how they might rate their speaking performance, their self-assessment result was used to compare with the scoring result averaged by the two raters. The results of the two groups were quite similar. Then, to test how they perceived their performance, they were asked to rate a five-point Likert scale on their achievement on the four grading speaking criteria. This perceived result corresponded to their self-assessment result which also

quite aligned with the raters' score. This finding again can prove that students will be able to excel in class with well-planned instruction where students can benefit from group practice, clear rubrics, and quality remarks. And this result has satisfied the objective of the study. The students' mean score of their self-assessment seems objective as they did not tolerate themselves in giving themselves a score.

However, this study owes some limitations during research conduct. The researchers wanted to employ two groups of students, one for the experimental to benefit the training of using the rubric to compare with the control group. But, due to having difficulty in finding voluntary students, the study has only one group. Another way to research this topic is to test their placement speaking performance prior to the employment of the rubric. This can be done to have data to compare with their outcome speaking performance at the end of the course. This can produce a more conceivable result. Although so, this present study is also able to highlight their self-assessment which can be said to closely match with the rating result by the examiners/raters. Finally, this study still cannot guarantee the students' adequate information about what they did for each grading criteria since the grading criteria were not broken into smaller parts. This issue reminds what Butler (2016) had mentioned when the author is still questioning about whether students can provide specific comments on grading criteria. This needs to be addressed in the future. The last point can be made here is to test their perceptions on the rubric itself. This is carried out to see if they completely agree with the rubric so more amendments can be implemented.

ACKNOWLEDGMENTS

REFERENCES

Ali, A.K. & Farhad, P. (2016). Place of grammar in English language teaching. *Language in India, 16*(2), 140-149.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation, 10*(3), 1-12.

Budasi, I. G., Agustini, K. & Risadi, Y. (2020). Powerpoint game, motivation, achievement: The impact and students' perception. *International Journal of Instruction, 13*(4), 509-522.https://doi.org/10.29333/iji.2020.13432a.

Butler, Y. G (2016). Self-Assessment of and for Young Learners' Foreign Language Learning. In M. Nikolov (ed.), *Assessing Young Learners of English: Global and Local Perspectives*, Educational Linguistics 25, http://dx.doi.org/10.1007/978-3-319-22422-0_12

Butler, Y. J. & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing 27*(1) 5–31. https://doi.org/10.1177/0265532209346370

Chien S.-Y., Hwang G.-J. & Siu-Yung Jong M. (2019). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education, 1*-48. https://doi.org/10.1016/ j.compedu.2019.103751.

Christiane, D. & Barbara, S. (1994). *Pronunciation.* Oxford University Press.

David, C. (2006). *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.

Guy, B-H. & Vanessa, J. (2012). *Complete IELTS Bands 5 -6.5, Student's book with answers*. Cambridge University Press.

Kim, J. (2019). Effects of Rubric-referenced self-assessment training on Korean high school students' English writing. *English Teaching, 74*(3), 79-11. https://doi.org/10.15858/engtea.74.3.201909.79.

Farooq, M., Ahmed, K. & Farooq, S. (2020). Introducing Self-Assessment for Evaluating Learners in Pakistan. *Global Social Sciences Review (GSSR)*, *5*(4), 120 – 136. http://dx.doi.org/10.31703/gssr.2020.

Hendra, H. (2012). Speaking problems faced by the English department students of Syiah Kuala University. *Lingua Didaktika, 6*(1), 37-44.

Huang, S. C. (2016). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education, 41*(6), 803–820, http://dx.doi.org/10.1080/02602938.2015.1042426

Mona, M. H. (2013). Factors negatively affect speaking skills at Saudi colleges for girls in the south. *English Language Teaching, 6(*12*),* 87-97. URL: http://dx.doi.org/10.5539/elt.v6n12p87.

NIVJA, H. (2011). Linguistic skills and speaking fluency in second language. *Applied Psycholinguistics*, 1-24. https://Doi: 10.1017/S0142716412000069.

Ou, Q. (2017). A brief introduction to perception. *Studies in Literature and Language, 15*(4), 18-28. http://dx.doi.org/10.3968/10055.

Pham, T. L. (2020). Perception of students at Bac Lieu University towards using interactive whiteboard in EFL classroom. *Journal of Language and Life,* 196-203. ISSN: 0868-3409.

Raja, M. I. K., Noor, R. M. R., Muhammad, S., Aino, H. I. & Ghulam, M. (2018). The role of vocabulary knowledge in speaking development of Saudi EFL learners. Arab World English Journal, 9(1), 406-418. https://dx.doi.org/10.24093/awej/vol9no1.28.

Reyad, O. A. R. &. Alhaj, A. A. (2020). The Importance of using speaking skill atsudanese EFL classroom. *International Journal of Multidisciplinary Research and Development, 7*(2), 95-102.

Seifert, T. & Feliks, O. (2018). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education,* 1-18. https://doi.org/10.1080/02602938.2018.1487023.

Shumei, Z. (2009). The role of input interaction and output in the development of oral fluency. *English Language Teaching*, *2*(4), 91-100.

Su, W. (2021). Exploring how rubric training influences students' assessment and awareness of interpreting. *Language Awareness, 29*(2), 178-196. https://doi.org/10.1080/09658416.2020.1743713.

Suryanti, N & Nurhuda (2021). The effect of problem-based learning with an analytical rubric on the development of students' critical thinking skills. *International Journal of Instruction, 14*(2), 665-684.

Thorben, J., Vogelin, C., Machts, N., Keller, S., Koller, O. & Moller, J. (2020). Judgment accuracy in experienced versus student teachers: Assessing essays in English as a foreign language. *Teaching and Teacher Education, 97*, 1-11.https://doi.org/10.1016.

Wang, W. (2016). Using rubrics in student self-assessment: student perceptions in the English as a foreign language writing context. *Assessment & Evaluation in Higher Education*, 1-14. http://dx.doi.org/10.1080/02602938.2016.1261993.

Wattie, M. (n.d.). S*peaking success book: Strategies and model answers*. https://www.answers.com/-Books.html.

APPENDIX

Rubric worksheet is adopted and adapted from Wattie (n.d.) including four grading speaking domains, using a 10 scale scoring system stipulated in Vietnam.

	1-2	3-4	5-6	7-8	9-10
Fluency & coherence	- speaks some individual words. - cannot make complete sentences. - shows too many hesitations.	-speaks at length, but some effort is required -shows loss of coherence due to repetition, hesitation, and self-correction - uses inappropriate linking words	- speaks at length without noticeable effort - doesn't show loss of coherence due to hesitation, correction or repetition - uses a wide range of linking words	- speaks fluently, hesitation and rarely to search for language - develops topics coherently and appropriately - uses a wide range of linking words and can make information highly coherent	- speaks like a native speaker, repetition or self- correction and hesitation (but not because of language difficulty; instead looks for ideas to speak) - uses a wide range of appropriate cohesive features
Pronunciation	- pronounces individual words not clearly. - cannot be understood.	- uses intonation, stress, and rhythm but with mixed control and it is not maintained. - can be understood throughout, though mispronunciation of individual words.	- uses intonation, stress, rhythm, and sustains flexible use of these features, with only occasionally leaps. -makes it easy to understand throughout and accent has minimal effect on intelligibility	- uses a wide range of pronunciation features flexibly and appropriately - makes it easy to understand throughout and accent has no effect on intelligibility	- uses a full range of pronunciation features with precision and subtlety. - sustains flexible use of pronunciation - can be understood effortlessly.
Grammatical range and accuracy	- uses some simple sentences but has incomplete and incorrect tense.	- uses a mix of simple and complex structures, but with limited flexibility -may make frequent mistakes with complex structures, though rarely cause comprehension problems.	- uses a wide range of complex structures - frequently produces error-free sentences	- uses a wide range of structures - shows that most sentences are error- free	- uses full of range of structures naturally and appropriately - produces consistently accurate structures.

Lexical resources	- has only some isolated words to speak. - makes so much effort to express his/her ideas.	- has sufficient vocabulary to discuss topics at length and make the meaning clear in spite of some errors - generally paraphrases successfully. - has insufficient advanced language.	- uses vocabulary flexibly to discuss a variety of topics. - paraphrases effectively. -uses some less common and idiomatic vocabulary, collocation with some appropriate choices.	- uses a wide range of vocabulary and flexibly convey precise meaning. - paraphrases effectively. - uses less common and idiomatic vocabulary skillfully.	- uses vocabulary with full flexibility and precision for all topics. - uses idiomatic language naturally and accurately.