

NAVIGATING THE PERCEPTION-PRACTICE DILEMMA: AN INQUIRY INTO ASSESSORS' IMPLEMENTATION OF ORAL PRESENTATION RUBRICS

Mahija Veetil, *Rashida Iqbal & Mohammed Abdulgalil Abugohar

Centre for Preparatory Studies, Dhofar University, Oman.

*Corresponding author: riqbal@du.edu.om

Received: 19.03.2024

Accepted: 26.06.2024

ABSTRACT

Background and Purpose: Assessing oral presentations involves evaluating multiple interconnected skills, including fluency, coherence, syntax, grammar, diction, and task achievement, making it a complex and challenging task. An oral presentation rubric serves as a tool to promote consistency and objectivity in such assessments. This study investigates the application and perceptions of the in-house speaking assessment rubric used by the Foundation Program (FP) English Unit at Dhofar University (DU) across levels 1, 2, and 3. The aim is to identify potential weaknesses and ambiguities in the rubric and provide recommendations to ensure fair, objective, and uniform assessments.

Methodology: This study employed a mixed-methods approach within an exploratory sequential design. Data were collected over three phases: a pre-session task (n=18), hands-on focus group sessions (n=16) to gather qualitative insights into the rubric's application, and a questionnaire (n=40) to obtain quantitative data on teachers' perceptions, cross-referenced with implementation practices.

Findings: The findings revealed significant gaps in the current rubric, leading to inconsistencies in grading practices for both formative and summative assessments. Despite assessors expressing confidence in the rubric's reliability and their level of training, notable discrepancies were observed between their perceptions and actual application of the rubric criteria. These inconsistencies highlight design flaws and ambiguities in the rubric, contributing to variability in assessment outcomes.

Contributions: This study underscores the necessity of regularly refining assessment rubrics to reduce ambiguity and subjective interpretation. It also emphasizes the importance of comprehensive user training to enhance assessment practices, support professional development, establish clearer learning

expectations for students, and minimize grading disparities. These findings offer valuable insights for improving the assessment of oral presentations in educational contexts.

Keywords: Rubric, oral presentation, assessment, discrepancy, raters.

Cite as: Veetil, M., Iqbal, R., & Abugohar, M. (2024). Navigating the perception-practice dilemma: An inquiry into assessors' implementation of oral presentation rubrics. *Journal of Nusantara Studies*, 9(2), 267-296. <http://dx.doi.org/10.24200/jonus.vol9iss2pp267-296>

1.0 INTRODUCTION

A rubric is a document that provides a description of performance using a varying scale with a set of criteria against which the student performance is measured (Aly, 2020). In English language teaching, assessment rubrics have always been reliable tools for consistent, objective and fair assessment of productive skills such as writing and speaking. While there are fewer grey areas in the assessment of reading and listening skills, productive skills such as writing and speaking are personal expressions and, therefore, are more complicated to assess than receptive skills because a degree of subjectivity always plays a role in the assessors' judgement of the quality of the production in writing and speaking. This subjective factor and the added challenge of assessing spoken production in real-time with limited or no possibility of reviewing complicate oral presentation assessment even more. Hence, a sound, fool-proof assessment tool is crucial for an objective, fair oral presentation assessment.

The effectiveness of an objective, fair assessment tool is a high priority (Ulker, 2017). A defective tool will result in a defective assessment, which directly impacts the student's grades and has larger repercussions for the other stakeholders – the teacher, the program, and the institution. Researchers agree that an oral presentation rubric must be well-defined to serve its purpose of a fair, objective, and wholesome assessment of oral production. Aiming at a fair and objective evaluation of students' performance, higher education institutions worldwide customarily include rubrics as an assessment tool. It is clearly evident that rubrics are quite effective tools for appraising various facets of oral presentations, including organization, content, delivery, and overall performance, providing clear guidance on evaluation criteria, assisting in shifting from a subjective evaluation to more objective and specific assessments (Belboukhaddaoui & Ginkel, 2019; Huang & Gui, 2014; Namin et al., 2021; Pérez-Torregrosa et al., 2022; Reddy & Andrade, 2010), and rubric criteria keep it simpler for learners to

understand what is required from them, and for the assessors to use it as their reference (Haji Suhaili & Haywood, 2017).

The Oman Academic Accreditation Authority (OAAA) stipulates the DU FP learning outcome standards. This stipulation further cascaded into the three levels of the FP, encompassing the scope and sequence progressing from a micro level to the macro level, eventually meeting up to the stipulated standard. The oral presentation rubric used in the Foundation Program is prepared in-house. It derives the criteria of assessment from the stipulations of QAAA standards with the rating scale matching the requirements of the IELTS speaking exam. This is because OAAA stipulates the IELTS 5.5 band as the required exit standard for the Foundation Program students in the Sultanate of Oman. The identical oral presentation rubric is used across the three levels following the progressive structure of the scope and sequence document. At DU FP, the in-house rubric has been used in oral assessments of students placed in levels 1, 2, and 3 for many years; however, it has been noticed that there arises a degree of discrepancy in grades during the double-blind marking. Hence, this research endeavours to examine the effectiveness and usability of the oral presentation rubric used.

2.0 LITERATURE REVIEW

Using rubrics in oral presentation assessments is a well-researched topic in educational settings. A lot of studies have highlighted the positive aspects of assessment rubrics and their essential role in ensuring objective scoring (Almohaimede, 2021; Christensen et al., 2011; Gunderson et al., 2021; Huang & Gui, 2014; Murillo-Zamorano & Montanero, 2018) while others could detect some drawbacks when putting rubrics in actual settings of evaluating oral presentations. An important study by Thala and Sheehan (2022) revealed that even if there is a standard rubric with set criteria, raters do affect the assessment process. The study deepened the understanding of the rating process by examining think-aloud protocols that 13 VSTEP-speaking raters supplied and using an analytical rating scale to evaluate 15 recorded performances with various proficiency levels. The information revealed that while all raters appeared to go through similar phases in the rating process, there were distinctions between the rating behaviours of inexperienced and seasoned raters. In oral presentation tests, raters are crucial to determine scores indicative of different aspects of performance; meanwhile considering their importance, they are also considered as one of the potential sources of bringing in undesired inconsistency in grades (Davis, 2012).

This has been further investigated by Duijm et al. (2018), who found that the rater's judgement affects student scores based on the raters' expertise and knowledge. It might also

influence in terms of leniency and considering different aspects of discourse. This study investigated two aspects of oral speech accuracy and fluency taking into consideration 68 stimuli in Dutch L2. Fifty-five professional raters and 41 non-trained, potential stakeholders holistically rated all stimuli. All raters had extensive prior exposure to spoken L2 Dutch. Linear mixed modelling showed that improvement of either fluency or accuracy led to significantly high value, considering that professional and nonprofessional raters had vast experience in L2 speech. The objective was to understand if they exhibit comparable responsiveness to these two elements. The study suggested that linguistically trained raters focused on accuracy more, considering the morpho-syntactical aspect, whereas the nonprofessional raters emphasized fluency.

Subjectivity in scoring and test bias are two issues raised concerning effectively assessing speaking, as Hidri (2018) asserted. In most cases, there is a chance that the speaking performance score will be unjust due to these two factors. The lives of test-takers and numerous other stakeholders may be harmed. The study examined the evaluation of speaking among English language learners in an EFL context inside an ESP program. Therefore, using a five-rubric scale (task achievement, fluency, grammar, vocabulary, and phonology), 12 raters evaluated the speaking abilities of 213 test-takers. Every exam question was formulated beforehand, and instructors were told to follow the exam's guidelines. Hidri's (2018) findings demonstrated that raters were prejudiced in favor of the speaking rubrics, indicating a hazy understanding of such rubrics and that test-takers speaking skills were typically assessed more leniently than harshly. Without a doubt, the several FACETS statistical tests demonstrated that the speaking exam lacked both validity and reliability. The study recommended creating a list of test requirements (specs) to create practical and equitable speaking exams, and it had significant implications for rater training on how to score speaking objectively.

In another study by Polat (2020), a Rasch analysis of rater behaviour conducted in 2019 on 6 expert graders to the performance of 24 English Language learners through video-recorded oral interviews confirmed that grader conduct during test taker oral skills assessment is necessary for both identifying the harsh and lenient graders in the rater group and acting accordingly to settle on the best matches for graders when paired rater-scorings or cross-marking-gradings are preferred for increasing the objectivity. A Rasch model in which MFRM (Many Faceted Rasch Measurement) explored the scoring performances of the expert graders in terms of rigorous and lenient scoring. The results revealed significant grade differences that might eventually affect learners' speaking grades significantly. Some raters scored too leniently or stringently, which might significantly affect learners' speaking grades.

Attempting to resolve the issue of oral or writing assessing discrepancy, some studies concluded that rater training (Mehmet & Karakaya, 2022) and group discussion among the raters (Ahmadi, 2020) could reduce rater subjectivity, rater dominance, and rater bias. Mehmet and Karakaya (2022) considered training to increase rater reliability and accuracy and decrease rater behaviour thus inspiring rater objectivity. Mehmet and Karakaya (2022) employed 45 raters, 23 from the control group and 22 from the experimental group using a quasi-experimental design. The data collection tools comprised a writing task developed by IELTS and an analytical rubric to assess academic writing skills. The study suggested that although rater behaviour affected test takers' scores, rater training effectively reduced rater errors and provided assessors with a fair and valid measurement in the performance assessment process. Likewise, Ahmedi (2020) applied group discussions to increase rater objectivity in speaking assessment, but in the process, was able to explore different shades of rater dominance and inconsistency in the grades. His study was conducted with five Iranian PhD students in a TEFL context. The study provided valuable insight into the complexity of rater dominance as a highly context-dependent issue and various factors such as rater personality, rater taking turns, rater bias, and understanding of the rubric. Informed by the literature, the current research came to add to those previous attempts with the speaking assessment rubric as the focus.

In conclusion, the literature indicated the effectiveness of rubrics in evaluating oral presentations as they provide clear assessment criteria and increase rater reliability. Still, it has not been quite effective in removing personal bias and perception to a fair level. Although rubrics provide a consistent assessment approach by defining performance criteria, guaranteeing that assessments are grounded in rules rather than personal judgments, the literature review indicates loopholes and encourages validation. The present study investigated this dilemma in the context of Omani higher education.

The current study provides pointers to ensure that the FP oral presentation rubric achieves fair, objective, and uniform marking in formative and summative assessments and to investigate the reasons behind discrepancies, if any, found. To this aim, the study was conducted to answer the following three questions:

1. Does the oral presentation rubric provide consistency in grading with different raters?
2. What are FP teachers' perceptions of the rubric implemented for assessing speaking skills?
3. Does inherent subjectivity influence the interpretation of the oral presentation rubric descriptors in academic assessment?

3.0 RESEARCH DESIGN

3.1 The Study Context

This study was conducted in the context of formative and summative assessment systems in the FP at DU. The oral presentation rubric was designed in-house based on the public version of the IELTS speaking band descriptors, as the FP program is geared towards the IELTS 5.5 band as the exit requirement. When some discrepancies in the speaking presentation grades were noted during the double-blind marking process during the routine assessment of the oral presentations in summative assessments, it was decided to evaluate the rubric and train the assessors who are the focus group of this research.

3.2 Data Collection Instruments

This mixed-methods study in an exploratory sequential design (Creswell & Plano Clark, 2011) commenced in January 2023. The study used qualitative and quantitative data from a focus group session and a questionnaire. The research was conducted incorporating qualitative and quantitative data across three phases.

In the first phase, the pre-session task, a sample (n=18) from the target population received a level 1 student's audio-recorded presentation via email, which they scored using the oral presentation rubric and entered the grades into a Google form. As discrepancies were observed in four key areas: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation, the pre-session task is focused on investigating these four criteria from the rubric. Two additional criteria, namely questions and length, were omitted. Since the criterion for length was assessed using a timer. At the same time, the grades for post-presentation questions had minimal impact on the overall assessment; these two criteria were excluded in the pre-session task.

For the second phase, the data collected was used in the focus group session (n=16) as a reference to aid in the understanding and application of the rubric to enhance its efficacy and to ensure a unified understanding regarding criteria and the descriptors outlined in the rubric. Teachers, divided into groups of four, watched a video-recorded student's oral presentation, and each group recorded their comments on the student's performance in light of the rubric criteria. A lot of valuable qualitative data was gathered from this phase.

The third phase, finally, gathered data on the assessors' perceptions of the rubric criteria and descriptors; a questionnaire was distributed electronically to all English language assessor teachers (n=45) in the Foundation Program who use the oral presentation rubric. The received

complete responses reached 40 at a completion rate of almost 89%. The questionnaire included two sections: Section A (7 questions) was to collect background about the respondents, while Section B (13 statements) targeted questions about subjectivity and difficulties encountered when using the rubric for achieving uniform grading. The results are presented in the data analysis section.

3.3 Validity & Reliability

First, the reliability statistics for the questionnaire (Section B) demonstrate a high level of internal consistency. With a Cronbach's Alpha coefficient of 0.939, the original and standardized item analyses show strong reliability indicating that the items within the instrument consistently measure the same underlying construct across the 13 items. Therefore, researchers can rely on the instrument to measure the intended variables in future studies. Second, the validity of the research instruments was assessed by two academicians and 5 professionals in the field of ELT. Their feedback was incorporated, enhancing the instruments' validity and credibility.

4.0 DATA ANALYSIS

4.1 Phase 1: Pre-Session Task Results

The pre-session task was administered preceding the focus group session, where eighteen teachers from the Foundation Program graded an oral presentation recorded by a level one student using the oral presentation rubric. The bar graph illustrates the grades assigned by the 18 evaluators to the Level 1 student's audio presentation, as in Figure 1.

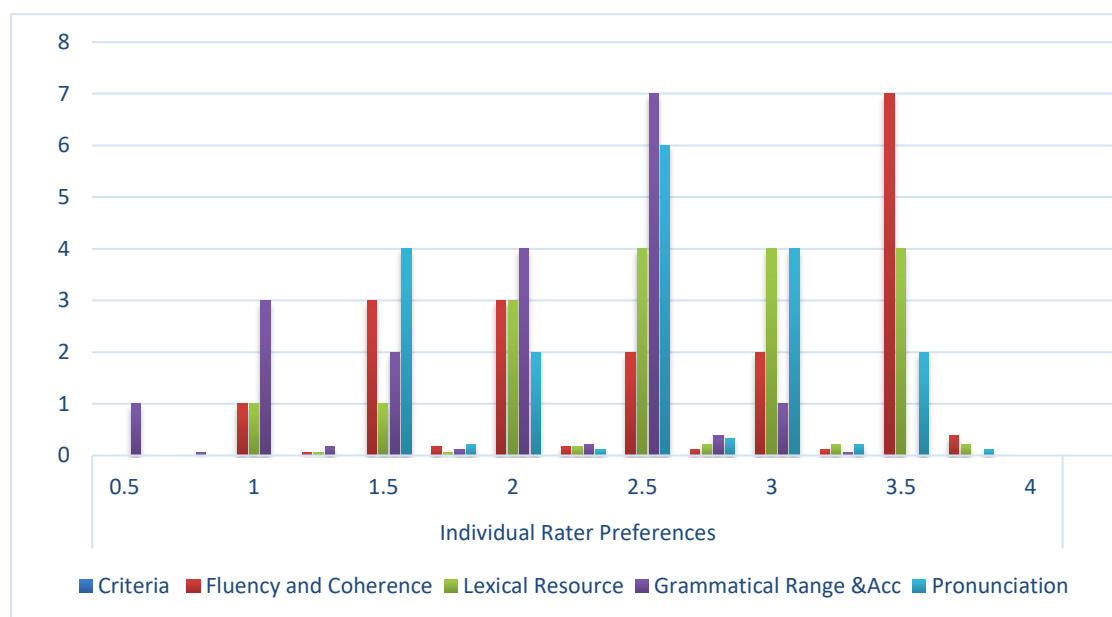


Figure 1: Pre-session raters' criteria marking

The range of scores suggests a significant variation in the assessed oral presentations. There is a difference of at least 2 marks in the grading of each of the criteria.

Table 1: Pre-session raters' marking - Total scores

Raters	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Total score/16	3.25	10	11.5	12	4	11	9.5	9.5	13.5	12.5	9	11.5	12	5.5	6	11.5	11.5	8

The highest and the lowest total scores awarded for the student's oral presentation are tabulated in Table 1. There is a difference of 10.25 marks between the lowest and the highest ratings, indicating an underlying issue within the rubric design or the rater behaviour.

4.2 Phase 2: Focus Groups Results

The session conducted for the focus groups, comprising teachers responsible for routinely assessing students' oral presentations using a rubric, provided valuable qualitative data. The rubric descriptors were broken into smaller segments and presented on a worksheet. Participants were organized into groups of four and assigned to evaluate a video presentation, conducted under exam conditions, by a level 1 student on the topic 'The History of a Business'. Each group focused on one of the four rubric criteria using the worksheet. Each group analyzed the presentation and marked segments that corresponded with their opinions. The gathered

responses yielded valuable insights into the assessors' evaluation of the student's presentation in relation to the rubric descriptors.

In this section, the rubric descriptors from the oral presentation rubric utilized in the Foundation Program are analyzed alongside the participants' observations of the student's video presentation recorded in the worksheet administered during the discussion session.

4.2.1 The rubric descriptors for fluency and coherence

The master rubric for fluency and coherence has outlined descriptors that classify examinees into 5 categories from unacceptable to excelling as in Table 2.

Table 2: Fluency and coherence descriptors

	20%
	Fluency & Coherence
<i>Excelling</i>	4 Excellent flow of speech with no reading; Very little repetition, pausing, or self-correction; Minor errors in complex communication; Full coherence; Exceeds expectations for this level
<i>Succeeding</i>	3 Good flow of speech with no reading; A little repetition, pausing, or self-correction; Minor incoherence ; Meets all expectations for this level
<i>Passing</i>	2 Adequate flow of speech with minimal reading; Some repetition, pausing, or self-correction; communication affected in places; Some incoherence; Meets most expectations for this level
<i>Emerging</i>	1 Limited flow of speech with reading; Repetition, pausing, or self-correction; communication affected frequently; Limited coherence; Meets some expectations for this level
<i>Unacceptable</i>	0.5 Little flow of speech or all reading; Little coherence; Meets very few expectations for this level

The participants' responses

Table 3: Application to fluency and coherence descriptors

Fluency	P1	P2	P3	P4
1. The student's pauses	1.2 non-lexical sounds present 1.3 Shows a state of anxiety. 1.4 Cognitive difficulty to recollect ideas	1.2 non-lexical sounds present 1.3 Shows a state of anxiety.	1.2 non-lexical sounds present 1.1 Shows a state of anxiety. 1.2 Cognitive difficulty to recollect ideas	NA

2. Student Referring to the script	2.1. should be penalized for lack of confidence.	2.2. could be overlooked if the flow is not interrupted	2.2. could be overlooked if the flow is not interrupted	NA
3. Repetition and Self-correction	3.3 disrupts fluency	NA	3.3. disrupts fluency 3.4. disrupts fluency and communication.	NA

As students' pauses, self-corrections, and reliance on scripts are significant components of the fluency descriptors, these aspects were divided into various segments (Table 3) and presented in the worksheet. All focus group participants who evaluated the fluency and coherence of the student's video presentation agreed that non-lexical sounds were present. However, Participant 2 (P2) did not perceive this as posing a cognitive difficulty in recollecting ideas. Similarly, one participant believed referencing the script should result in penalties, while others suggested it could be overlooked. These differing opinions suggest that these participants will likely assign different fluency grades. For instance, those who overlook script referencing may grade the student higher, while the one who advocates for penalties might grade harshly. Concerning repetition and correction, only one participant considered it disruptive to communication. Consequently, this participant is more inclined to grade the student lower, possibly as 'passing' (=2/3) or 'emerging' (=1/3), as specified in the rubric descriptors for these rating scales.

4.2.2 The rubric descriptors for lexical resources

The master rubric for lexical resources has outlined descriptors that classify examinees into five categories, from unacceptable to excelling, as in Table 4.

Table 4: Lexical resources descriptors

20%
Lexical Resource
4
Accurate, appropriate range of taught vocabulary; Minor errors in complex or creative lexis; communication not affected Exceeds expectations for this level
3
Appropriate range of taught vocabulary; A few errors; communication not affected; Meets all expectations for this level
2
Adequate range of taught vocabulary; Some errors; communication affected in places; Meets most expectations for this level
1
Limited range of taught vocabulary; Many errors; communication affected frequently; Meets some expectations for this level
0.5
Little range of taught vocabulary; Rife with errors; little communication; Meets very few expectations for this level

The participants' responses

Table 5: Application to lexical resources descriptors

Lexical Resources	P1	P2	P3	P4
4.1 Business collocation	was founded Company was expanded	Company expanded	Founded by Verb be+ located	Opened a branch
4.2 Business vocabulary	Branch, employ. Employees	Founded, open, located, branch, employees	Founded, expanded, employ, employees, branches, celebrate, strong policy, areas, program	Expanded, opened, branch, located, services
4.3 Comments	expanded pronounced incorrectly	NA	Not many collocations, Good range of target vocabulary, appropriate not affected	NA

4.4. Opinion about the usage	Basic vocabulary choice, Lapses are noticeable, but minimal strain for the listener, Many errors – difficult to understand, target vocabulary not evident or incomprehensible	Basic vocabulary choice	A wide range of well-chosen business vocabulary	Basic vocabulary choice, Lapses are noticeable, but minimal strain for the listener, Many errors – difficult to understand, target vocabulary not evident or incomprehensible
------------------------------	---	-------------------------	---	---

The rubric descriptors for this criterion mainly focus on the student's use of taught vocabulary and complex creative lexis (Table 5). The vocabulary taught here mainly consists of business vocabulary and business collocations. Participants of this group who focused on lexical resources were asked to jot down the business words and collocations used by the student in the task sheet while watching the student's video presentation. While one participant wrote down almost nine words, others could only capture about five. The task is understandably challenging to transcribe while watching. However, it is interesting to analyze the participants' opinions on the student's vocabulary usage. The participant who listed the most business words noted a lack of collocations but praised the wide range of business vocabulary. In contrast, others observed the use of only basic vocabulary, the absence of target vocabulary, or incomprehensible language. This suggests that their evaluations and interpretations differed, suggesting differing perceptions of the adequacy and effectiveness of the student's language use. This could stem from either a lack of attentiveness or a subjective assessment of the quality and quantity of vocabulary employed by the student – both a common occurrence during the assessment of oral presentations in classrooms. The differences in grading can be attributed to both these factors playing a role in assessment scenarios.

4.2.3 The rubric descriptors for grammatical range and accuracy

The master rubric for grammatical range and accuracy has outlined descriptors that classify examinees into 5 categories from unacceptable to excelling as in Table 6.

Table 6: Grammatical range and accuracy descriptors

20%
Grammatical Range & Accuracy
4
Excellent range of taught structures; Minor errors; communication not affected; Exceeds expectations for this level
3
Good range of taught structures; A few errors; communication not affected; Meets all expectations for this level
2
Adequate range of taught structures; Some errors; communication affected in places; Meets most expectations for this level
1
Limited range of taught structures; Many errors; communication affected frequently; Meets some expectations for this level
0.5
Little range of taught structures; Rife with errors; little communication; Meets very few expectations for this level

The participants' responses

Table 7: Application to grammatical range and accuracy descriptors

Grammar	P1	P2	P3	P4
5.1 Past tense verb forms	Some errors	Many errors	Some errors	Some errors
5.2 Structure and Time Phrases /Clauses	Mostly accurate Lack clarity	Mostly accurate, time phrases used only once	Some errors	Mostly accurate
5.3 Subject-Verb Agreement	Some errors	Some errors	Some errors	Mostly accurate

As evident in Table 7 from the responses, it is interesting to compare the observations of Participants 3 and 4. Participant 4 believes that the taught grammar structures were predominantly used accurately by the student, whereas Participant 3 pointed out some errors. participant 1 remarked that while the time phrases and clauses were mostly accurate, they lacked clarity. The differing perspectives should be examined in light of the significant grade variation observed for this criterion in the pre-session task. Intriguingly, these responses are not strikingly indicative of such a major discrepancy in grades.

4.2.4 The rubric descriptors for pronunciation

The master rubric for pronunciation has outlined descriptors that classify examinees into 5 categories, from unacceptable to excelling, as in Table 8.

Table 8: Pronunciation descriptors

20%
Pronunciation
4
Excellent pronunciation; Effortless to understand; Minor impact from L1; Exceeds expectations for this level
3
Good pronunciation; Easy to understand; Some impact from L1; communication not affected; Meets all expectations for this level
2
Adequate pronunciation; Mostly easy to understand; Some impact from L1; communication affected in places; Meets most expectations for this level
1
Limited pronunciation; Difficult to understand in places; Frequent impact from L1; communication affected frequently; Meets some expectations for this level
0.5
Poor pronunciation; Difficult to understand throughout; Pervasive impact from L1; little communication; Meets very few expectations for this level

The participants' responses

Table 9: Application to pronunciation descriptors

Pronunciation	P1	P2	P3	P4
6.1. Above A2 * CEFR requirement (corresponds to basic users of the language)	Good enough	NA	NA	The language is clear and easy to understand, minor pronunciation errors
6.2. Easily understandable some errors	Mostly understandable	NA	Easily understandable – some errors	NA
6.3. Understandable to some extent – some words are not clear	NA	Can understand – some words stressed more	NA	NA
6.4. Hard to understand – heavy L1 impact	NA	NA	L1 impact	NA

Based on the responses in Table 9, most participants who focused on the pronunciation criteria emphasized that students' presentation was generally clear and easily understandable, with pronunciation not hindering clarity. Only one participant noted the presence of L1 impact,

despite its significant inclusion in the rubric. This highlights the assessors' tendency not to consider L1 impact as a scoring factor, despite the descriptors emphasizing its importance.

4.3 Phase 3: Data Analysis of The Questionnaire

In this analysis, we examine the responses of 40 participants who provided feedback on the oral presentations in the Foundation Program.

The survey questionnaire consisted of two sections. Section A aimed to collect data about the respondents' teaching profile and engagement level with the oral presentation rubric used in the Foundation Program at Dhofar University. Section B targeted the respondents' perceptions of the rubric and its components. This dual-section approach comprehensively captures the assessors' demographic profile and their nuanced perspectives on the rubric's components, providing a holistic understanding of the assessment process within the Foundation Program.

4.3.1 Section A: Demographic data

The data in Figure 2 illustrates the composition of the respondents (n=40) based on the position of the respondent, including Teaching Assistant, Instructor, Lecturer, and Assistant Professor. The vast majority of the respondents were instructors and lecturers.

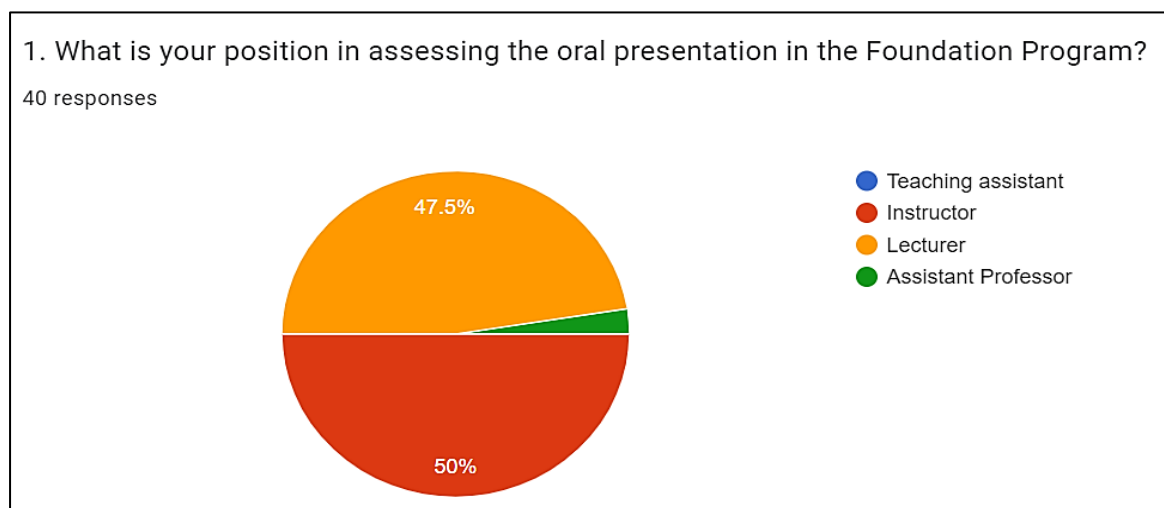
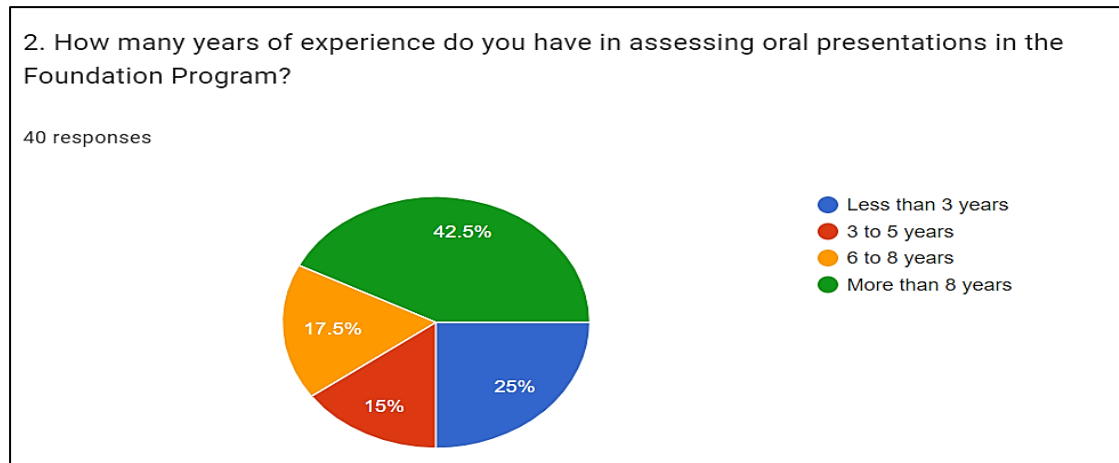


Figure 2: Section A Q1

The second question sought to understand respondents' experience levels in assessing oral presentations within the Foundation Program (Figure 3). A significant portion of respondents (42%) possess more than 8 years of experience assessing oral presentations. The 6 to 8 years and 3 to 5 years categories comprise 17.5% and 15% of respondents, respectively. A quarter of

respondents (25%) reported having less than 3 years of experience in assessing oral presentations.

Figure 3: Section A Q2



The next question enquired about the specific levels within the Foundation Program where respondents assess oral presentations. Most respondents, 82.5%, indicated that they assess oral presentations across all three levels of the Foundation Program (Figure 4).

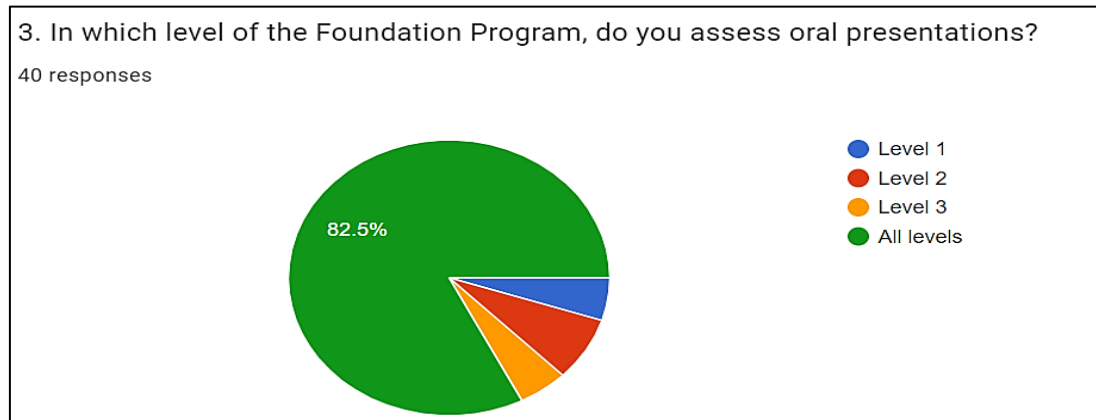


Figure 4: Section A Q3

The fourth question concerns the assessors' familiarity with the FP learning outcomes specifically related to oral presentations. The responses reflect varying levels of engagement with the learning outcomes. Half of the group (50%) reported being fully aware of the GFP learning outcomes of oral presentations. A little below that figure, 42.5%, claimed to have read the learning outcomes thoroughly. A smaller proportion, 7.5%, admitted to having only glanced

at the learning outcomes, and none for 'Not at all', indicating a positive trend of engagement and awareness among the assessors (Figure 5).

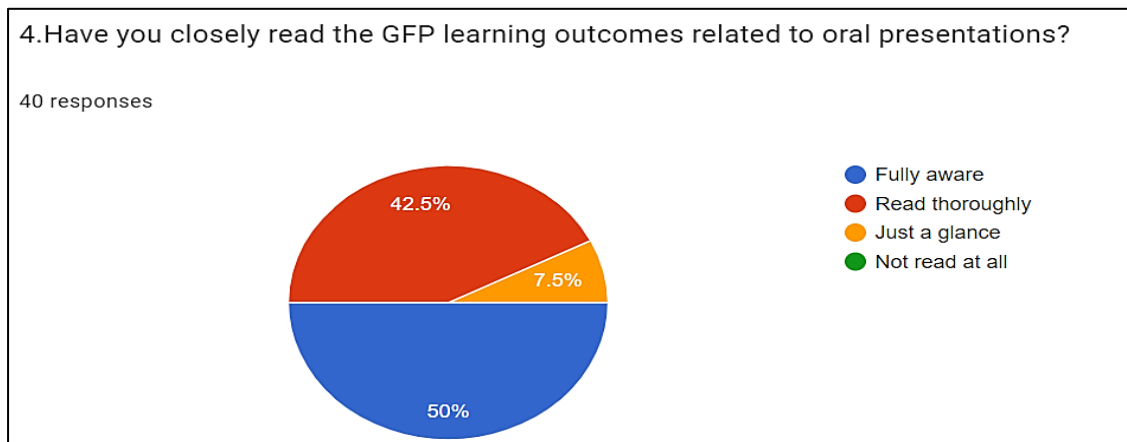


Figure 5: Section A Q4

The following question aimed to assess the extent to which assessors utilize the current oral presentation rubric for Formative Assessment within the Foundation Program. A vast majority, comprising 95% of respondents, use the current oral presentation rubric for formative assessment. A small percentage (5%) indicated that they did not use the current rubric for formative assessment (Figure 6).

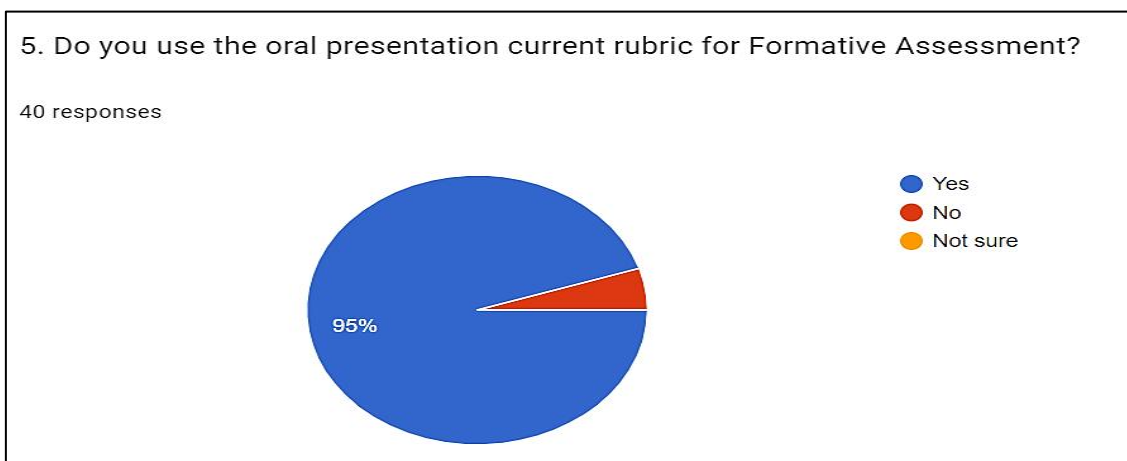


Figure 6: Section A Q5

The figures were similar for the question regarding using the oral presentation rubric specifically for summative assessment within the Foundation Program. Exactly 95% of respondents reported using the current oral presentation rubric for summative assessment. A

small percentage (5%) indicated that the current rubric was not used for summative assessment (Figure 7).

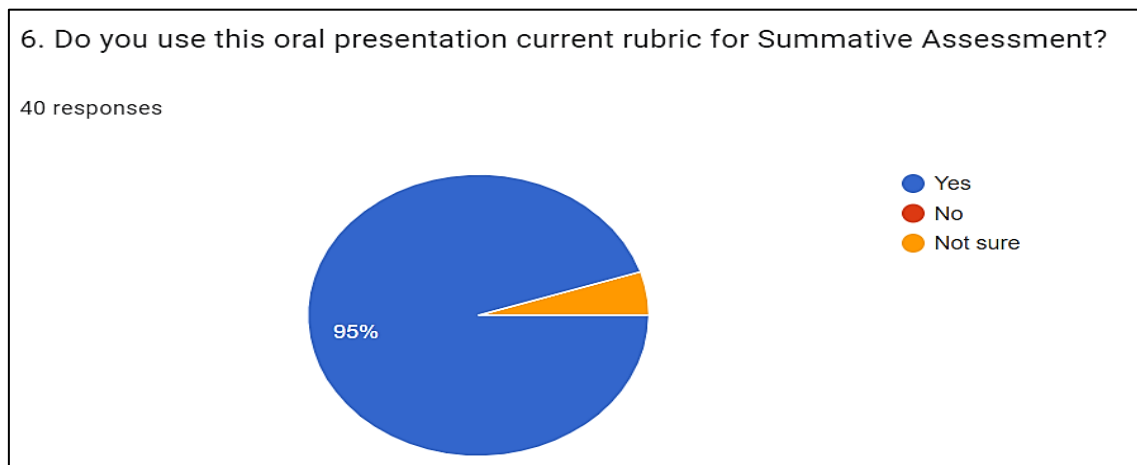


Figure 7: Section A Q6

The last question aims to evaluate assessors' satisfaction with the training they have received till the date of data collection regarding the utilization of the current rubric in their assessments within the Foundation Program. As in Figure 8, the majority (72.5%) reported that they had sufficient training, and together with those who feel highly trained (17.5%), they represent 90% of participants; however, only 10% are not satisfied with the training they received.

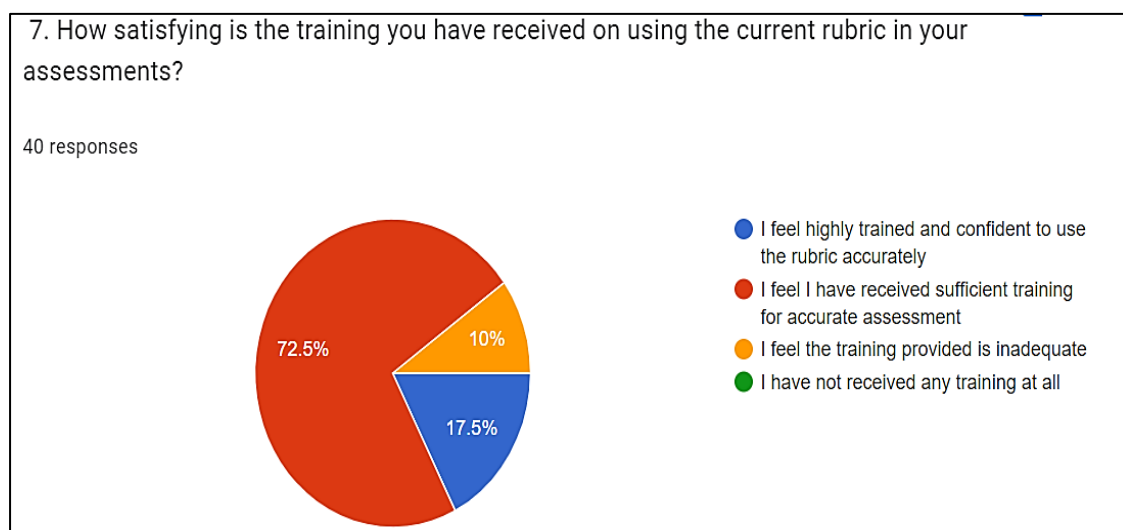


Figure 8: Section A Q7

4.3.2 Section B: Perceptions questionnaire results

Table 10: Perceptions' results (S1-S13) descriptive statistics

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
N	Valid	40	40	40	40	40	40	40	40	40	40	40	40	40
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean		3.70	3.83	3.38	3.68	3.33	3.35	3.10	3.35	3.68	3.55	3.40	4.00	3.53
Median		4.00	4.00	4.00	4.00	3.50	3.00	3.00	4.00	4.00	4.00	4.00	4.00	4.00
Std. Deviation		.992	.903	1.213	.888	.944	.921	1.033	.975	.888	.846	1.033	.641	.847
Variance		.985	.815	1.471	.789	.892	.849	1.067	.951	.789	.715	1.067	.410	.717
Skewness		-.341	-.738	-.418	-.447	-.137	-.361	.232	-.427	-.678	-.834	-.159	-.616	-.749
Std. Error of Skewness		.374	.374	.374	.374	.374	.374	.374	.374	.374	.374	.374	.374	.374
Kurtosis		-.836	.046	-.823	-.366	-1.061	-.101	-.730	-.513	-.165	-.275	-1.211	1.730	-.392
Std. Error of Kurtosis		.733	.733	.733	.733	.733	.733	.733	.733	.733	.733	.733	.733	.733
Range		3	3	4	3	3	4	4	4	3	3	3	3	3
Minimum		2	2	1	2	2	1	1	1	2	2	2	2	2
Maximum		5	5	5	5	5	5	5	5	5	5	5	5	5
Sum		148	153	135	147	133	134	124	134	147	142	136	160	141

The questionnaire's Section B Statements (n=13) were to collect respondents' perceptions of the oral assessment rubric. Table 10 tabulates all descriptive statistics to the results collected for the 13 statements. It presents the mean scores, standard deviation, and range, among other calculations. Statements 1 to 13 (S1-S13) can be further classified into three categories: the rubric's general features, skills descriptors, and other criteria. Herein, collected results are analyzed along with the rubric item rating analysis.

A| The Rubric's general features

Statement 1 (S1) to Statement 7 (S7) collected assessors' evaluation of the general features of the oral presentation assessment rubric. Table 11 depicts the results collected for these seven statements.

Table 11: Rubric's general features results (n=40)

	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
	F*	%	F	%	F	%	F	%	F	%
S1	0	0	6	15	9	22.5	16	40	9	22.5
S2	0	0	5	12.5	5	12.5	22	55	8	20
S3	3	7.5	8	20	7	17.5	15	37.5	7	17.5
S4	0	0	5	12.5	9	22.5	20	50	6	15
S5	0	0	10	25	10	25	17	42.5	3	7.5
S6	1	2.5	6	15	14	35	16	40	3	7.5
S7	1	2.5	12	30	13	32.5	10	25	4	10

- F = Frequency, % = percent, S = Statement

The first statement (S1) said, *'The oral presentation rubric currently used in the Foundation program clearly reflects the objectives targeted for the assessment.'* Investigating deeper into the respondents' engagement with the rubric in the 2nd section of the survey, this question seeks to gauge the assessors' perceptions regarding the alignment of the oral presentation rubric with the intended assessment objectives within the Foundation Program. The responses to Q1 came at the mean score of 3.70 (SD 0.992). A combined total of 62.5% (22.5% strongly agree + 40% agree) of respondents expressed a positive view, indicating that they believe the rubric effectively reflects the targeted assessment objectives. A notable 22.5% maintain a neutral position on aligning the rubric with assessment objectives, suggesting a segment of assessors who may require further clarification or have mixed perceptions. A relatively smaller percentage, 15%, disagrees with the assertion that the rubric aligns well with the assessment objectives (Table 11).

S2: *'The 6 criteria (i.e., Fluency & Coherence, Lexical resources, Grammatical Range and Accuracy, Pronunciation, Length, and Discussion Questions) in the oral presentation rubric are inclusive enough to assess the attainment of all the LOs set for each course level.'* This statement (S2) seeks to evaluate assessors' understanding of the comprehensiveness of the six criteria within the oral presentation rubric in assessing the achievement of Learning Outcomes (LOs) across different course levels within the Foundation Program. The mean score of the responses is 3.83(SD 0.903). A majority of 75% (20% strongly agree + 55% agree) hold positive views, indicating a consensus that the six criteria effectively encompass the assessment of all LOs for each course level. A combined 25% (12.5% disagree + 12.5% neutral) expresses

dissent or neutrality. Notably, no respondents strongly disagree with the assertion that the criteria adequately cover the assessment of all Los (Table 11).

S3: *'The 5 grading scales (i.e., Unacceptable, Emerging, Passing, Succeeding, and Excelling) in the oral presentation rubric truly differentiate the various levels of students' performance.'* This statement aims to assess assessors' perspectives on the effectiveness of the five grading scales within the oral presentation rubric in discerning the different levels of student performance. The mean score is 3.38 (SD 1.213). For S3 in Table 11, a combined 55% (17.5% strongly agree + 37.5% agree) affirm that the grading scales effectively distinguish students' performance across various levels. 37.5% (17.5% neutral + 20% disagree) hold a neutral or dissenting stance. A small percentage, 7.5%, strongly disagrees with the statement that grading scales effectively differentiate student performance.

S4: *'The rubric helps the marker to grade the speaking presentation objectively to a great extent.'* This statement evaluates assessors' perceptions of how well the rubric objectively assesses the oral presentation. The mean score is 3.68 (SD 0.888). The data in Table 11 reveals a majority consensus of 65% (50% agree + 15% strongly agree) among assessors, signifying that most of the assessors rely on the rubric as an aid to grade the oral presentation objectively. The remaining 35%, out of which 12.5 %, dispute the objectivity of the rubrics grading criteria, suggesting a need for further exploration into potential areas of concern.

S5: (Table 11): *'The oral presentation rubric provides consistent results.'* Evaluating the consistency of results obtained through the oral presentation rubric, the data presents a diverse range of perspectives among respondents: a relatively small percentage strongly agrees that the rubric provides consistent results. The mean score is 3.33 (SD 0.944). A relatively small 7.5% strongly agree that the rubric provides consistent results. A quarter of the respondents express disagreement with the statement, indicating a notable portion that perceives inconsistency in the results generated by the rubric. An equal percentage of respondents remain neutral on the issue of the rubric's consistency. This suggests a significant number of assessors who neither affirm nor dispute the tool's ability to yield consistent results. The largest percentage falls in the agree category, 42.5%, indicating that most respondents believe the oral presentation rubric provides consistent results.

S6: *'The oral presentation rubric assists different markers to arrive at convergence of scores for the same student.'* For the statement, the oral presentation rubric assists different markers in arriving at convergence of scores for the same student; the mean score of the responses is 3.35 (SD 0.921). A total of 17.5% (n=7) either strongly disagree or disagree with the statement, indicating that some respondents believe the rubric is ineffective in facilitating

convergence of scores among different markers for the same student. A significant portion, 35%, adopts a neutral position on the rubric's effectiveness in achieving score convergence. A substantial majority of 40% (n=16) agree with the statement, indicating a positive perception that the oral presentation rubric is effective in assisting different markers to converge on scores for the same student. A smaller but noteworthy percentage of 7.5% strongly agree, suggesting a subset of respondents who strongly believe in the rubric's efficacy in achieving convergence of scores among markers (see Table 11).

S7: *'The descriptors for the various criteria in the oral presentation rubric are clear, leaving no room for ambiguity.'* The data on the clarity of descriptors in the oral presentation rubric indicates various opinions among respondents. The mean score is 3.10 (SD 0.033). A small proportion of 10% strongly agrees that the descriptors for the various criteria in the rubric are clear, suggesting a strong affirmation of clarity. A quarter of the respondents agree, indicating a larger portion with a positive view on the clarity of descriptors in the rubric. Table 11 shows that a significant number, 32.5%, adopts a neutral stance, suggesting a substantial number of respondents who neither strongly affirm nor dispute the clarity of descriptors. 30% disagree, signalling a notable dissent regarding the clarity of descriptors in the rubric. A negligible small percentage of 2.5% strongly disagrees, representing a minority.

B| The Rubric's skills descriptors

Statement 8 (S8) to Statement 11 (S11) collected responses related to four skills' descriptors, namely for Fluency and Coherence, Lexical Resources, Grammatical Range & Accuracy, and Pronunciation as shown in Table 12.

Table 12: Skills descriptors results (n=40)

	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
	F	%	F	%	F	%	F	%	F	%
S8	1	2.5	8	20	10	25	18	45	3	7.5
S9	0	0	6	15	6	15	23	57.5	5	12.5
S10	0	0	7	17.5	6	15	25	62.5	2	5
S11	0	0	11	27.5	7	17.5	17	42.5	5	12.5

S8: (Table 12).: *'The descriptors for 'Fluency & Coherence' are inclusive and task-based with all assessment elements in.'* The data on the descriptors for 'Fluency & Coherence' in the oral presentation rubric indicates varied opinions among respondents. The mean score is 3.35 (SD

0.975). A small percentage of 7.5% strongly agrees that the descriptors for 'Fluency & Coherence' are inclusive and task-based, indicating a strongly positive view regarding the assessment elements. A significant majority of 45% agree, suggesting a larger portion with a positive perception of the inclusiveness and task-based nature of the descriptors. A quarter of the respondents adopt a neutral stance, indicating a substantial number of individuals who neither strongly affirm nor dispute the inclusiveness and task-based nature of the descriptors. A notable portion of 20% disagrees, signalling a dissenting view on the inclusiveness and task-based nature of the descriptors for 'Fluency & Coherence'. A small percentage strongly disagrees, representing a minority with a particularly strong negative view of the inclusiveness and task-based nature of the descriptors.

S9: *'The descriptors for 'Lexical Resources' thoroughly assess the target vocabulary, appropriate register, effective collocations, and use of signposts within the context of the presentation.'* For the descriptors related to 'Lexical Resources' in the oral presentation rubric, the data provides insights into assessors' opinions. The mean score is 3.68 (SD 0.888). 12.5% strongly agree that the descriptors for 'Lexical Resources' thoroughly assess the target vocabulary, appropriate register, effective collocations, and use of signposts within the presentation context. This indicates a positive and strong affirmation. Most respondents, comprising 57.5%, agree with the statement. More than half of the respondents have endorsed the descriptors as effective in assessing the specified lexical elements within the presentation context. A moderate percentage, 15%, maintains a neutral stance on the descriptors for 'Lexical Resources' indicating uncertainty. Similarly, 15% disagree, questioning the descriptors' effectiveness (Table 12).

S10: *'The descriptors for 'Grammatical Range & Accuracy' practically aid in thoroughly assessing the intelligibility of verb tense usage, sentence structures, and level-specific target structures.'* The data related to the descriptors for 'Grammatical Range & Accuracy' in the oral presentation rubric provides insights into assessors' opinions. The mean score is 3.55 (SD 0.846). 62.5% agree that the descriptors for 'Grammatical Range & Accuracy' practically aid in thoroughly assessing intelligibility, verb tense usage, sentence structures, and level-specific target structures. A moderate 15% maintains a neutral stance, while 17.5% disagrees, opposing the practical effectiveness of the descriptors. Interestingly, 5% strongly agree that the descriptors practically aid thorough assessments (see Table 12).

S11: *'The descriptors for 'Pronunciation' consider the specific pronunciation features and challenges typically encountered by students of the Foundation Program.'* The data relating to the descriptors for 'Pronunciation' in the oral presentation rubric offers insights into

assessors' opinions, as in Table 12. The mean score is 3.40 (SD 1.033). As evident in Table 21, 42.5% agree, and 12.5% strongly agree the descriptors for 'Pronunciation' take into account specific pronunciation features and challenges typically encountered by students of the Foundation Program, indicating a positive perception among assessors regarding the relevance of the descriptors. A considerable 27.5% were sceptical of the effectiveness of the descriptors in addressing students' specific pronunciation challenges. A substantial minority, 17.5%, maintains a neutral stance on the adequacy of the descriptors to address the specific pronunciation features and challenges.

C] The Rubric's other criteria

Two statements (S12 & S13) are the last two in Section B of the survey. S12 is on the prescribed presentation length, while S13 is for post-presentation questions (see Table 13).

Table 13: Rubric's other criteria results (n=40)

	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
	F	%	F	%	F	%	F	%	F	%
S12	0	0	1	2.5	5	12.5	27	67.5	7	17.5
S13	0	0	7	17.5	7	17.5	24	60	2	5

S12: *'The prescribed length specified in the oral presentation Rubric is consistent with the expected requirements of the assessed level.'* The data related to the prescribed length specified in the oral presentation rubric revealed the assessors' perspectives very clearly. The mean score is 4.00, and the standard deviation is 0.641. About a fifth (12.5%) strongly agree that the length specified in the oral presentation rubric is consistent with the expected requirements of the assessed level, while more than half (67.5%) agree. A small percentage, about the eighth (12.5%), remained neutral on the appropriateness of the specified length. On the contrary, a very small percentage, 2.5%, strongly disagrees, expressing a strong negative view on the consistency of the prescribed length with the expected requirements (Table 13).

S13: *'The descriptors of the post-presentation discussion ensure consistent and standardized grading of the oral responses.'* The data related to the descriptors of the post-presentation discussion in the oral presentation yielded a mean score of 3.53 (SD 0.847), shedding light on the assessors' perception of this criteria. A significant majority (60%) agree that the descriptors of the post-presentation discussion ensure consistent and standardized grading of oral responses, whereas 17.5% maintained a neutral stance. An equal percentage of

17.5% disagrees, disputing the effectiveness of the descriptors in ensuring consistent and standardized grading. A small percentage of 5% strongly agrees that the descriptors contribute significantly to the consistency and standardization of grading (see Table 13).

5.0 DISCUSSION

In the first phase, analyzing the grades awarded for the level 1 student's audio presentation by the different graders, it is observed that the lowest range of scores indicates that some graders have rated the presentation range from below passing or unacceptable on the rating scale of the rubric. Contrastingly, the upper range of scores denotes some graders rated the presentation as outstanding. This disparity is surprising, pointing to a defect that could either be attributed to the loopholes in the rubric design or arising from the inherent subjectivity from the side of the graders introducing an element of inconsistency in the assessment process or a gap in training in rubric utilization.

In the second phase, the qualitative data gathered from the focus group session, where each group of participants concentrated on a specific rubric criterion while assessing a student's video presentation under exam conditions, reveals a notable disparity in assessment. The origins of the disparity intriguingly raise suspicions of subjective factors influencing the assessors' judgments, potentially interfering with the evaluation process. Further investigation into these factors is warranted to better understand and address their impact on assessment outcomes. It is equally surprising to observe that the criterion of grammatical accuracy, which exhibited the greatest disparity in grades in the pre-session task of assessing the student's audio presentation, did not emerge as a significant discrepancy in the observations and comments made by the assessors during the evaluation of the video presentation by the student.

In the third phase, the quantitative data collected on the assessors' perspectives of the oral presentation rubric revealed that while most of the respondents hold a positive view of the rubric's alignment with assessment objectives and the comprehensiveness of the criteria within the oral presentation rubric, the presence of neutral and dissenting perspectives highlights the need for further investigation and potential refinement of the rubric. The neutral and disagreeing responses, although minor, reflect a need to improve the comprehensiveness of the criteria and resolve ambiguities to fully align the rubric with the assessment. On the other hand, there is a mixed response regarding the five grading criteria in the oral presentation rubric in grading the level of performance. Although almost half of the respondents hold a positive view, more than a quarter are either neutral or disagree. This level of uncertainty or dissent is quite significant. The minority who strongly disagree with the effectiveness of the grading criteria

in differentiating the various levels of performance points to a need for potential refinement of the criteria or the descriptors of the criteria.

The data illustrates a significant majority consensus, with 65% of assessors (50% agree + 15% strongly agree) indicating their reliance on the rubric as a tool to objectively grade oral presentations. This suggests that most assessors accept the rubric's effectiveness in guiding assessment practices. However, it's notable that 35% of assessors, including 12.5% who dispute the objectivity of the rubric's grading criteria, underscore indicate some level of apprehension. This gives rise to a need to review the efficacy of the rubric in aiding objective assessment and to redefine the descriptors to resolve ambiguity. Almost half the respondents believe that the oral presentation rubric provides consistent results. However, the remaining 50% dispute the rubric's ability to produce consistent results. There is a divided perspective among respondents and therefore requires improvement to enhance the reliability of the rubric.

Results showed that more than half (52.5 %, S6, Table 16) are unsatisfied with the rubric and are sceptical about the rubric yielding uniform grades. It is evident that there is almost 50-50 division in assessors' views on the consistency and uniformity of grades generated by the rubric. Furthermore, 65% of the assessors (S7, Table 17) are critical of the clarity of the rubric descriptor out of which a sizable 32% are neutral expressing a level of uncertainty. This discrepancy is reflected in the significant variance in grades awarded by different graders for students' audio presentations. Corroborating the grades with the respondents' perspectives points to where the defect lies. These questions can serve as guiding factors to effectively address the identified issues.

Upon examining the assessors' perspectives on each criterion individually, it's evident that, except for fluency and coherence, where responses were evenly divided, the remaining criteria leaned more towards an agreement, with a ratio of approximately 70 to 30. However, the agreement was slightly lower for grammar at 62.5%, with 32.5% expressing neutrality or dissent. Overall, most assessors demonstrate trust in the reliability of the criteria descriptors, although the dissenting minority voices warrant further reflection. Corroborating this generally positive outlook with a consistent gap of at least 2 marks observed across all criteria in the grading of the recorded presentation in the first phase, a major discrepancy is evident, particularly considering that the highest band score for each criterion is 4. Therefore, this difference holds significant weight.

The data also highlights an interesting phenomenon - a gap between perception and application. The failure to arrive at a convergence of grades using the oral presentation rubric, analyzed from various perspectives using quantitative and qualitative data, appears to unveil a

mixed outcome. Although the data reveals the assessors' confidence and trust in the reliability of the rubric and the adequacy of training received, there seems to be a great disparity in grades, indicating a mismatch in perception vs application. This raises the question of subjectivity—whether the rubric permits subjectivity to influence the assessment of oral presentations. This discrepancy warrants further investigation. This is a gap that needs to be investigated.

The current study's findings were in the same vein as the studies that indicated that rater judgement, expertise and knowledge do affect the rating process even after implementing an analytical rubric or narrowing the investigation to some specific criteria of oral speech, such as accuracy and fluency, discrepancies in rater grades are quite obvious (Duijm et al., 2018; Thala & Sheehan, 2022). This similarity may be due to what Hidri (2018) concluded after incorporating a five-rubric scale with clear criteria for assessment and examining 213 test takers. Still, the results were quite surprising as they indicated harsh and lenient marking even after providing exam questions and guidelines ahead of time (Hidri, 2018). This strengthens the notion that no matter how well the rubrics are formulated, factors like hazy understanding of the rubric and personal perception affect the grades. The examiners cannot detach their personal preferences and understanding of the given criteria in the rubric, which leads to inconsistent grading. Moreover, further investigation into cross-marking grading and paired rater scoring also revealed significant grade differences (Polat, 2020). In addition, even measures like group discussions or think aloud protocols could not bring the raters on one page (Ahmadi, 2020; Mehmet & Karakaya, 2022). On the contrary, the findings here came in contrast to some earlier studies that discussed rubrics as a fair documents (Aly, 2020) for effective assessment (Ulker, 2017), providing clear assistance on evaluation criteria and focusing on objectivity (Almohaimede, 2021; Christensen et al., 2011; Gunderson et al., 2021; Huang & Gui, 2014; Murillo-Zamorano & Montanero, 2018) and specific evaluation instead of subjective assessment (Belboukhaddaoui & Ginkel, 2019; Huang & Gui, 2014; Namin et al., 2021; Pérez-Torregrosa et al., 2022; Reddy & Andrade, 2010).

6.0 CONCLUSION & RECOMMENDATIONS

Throughout this research, the endeavour is to help the rubric setter identify and plug the gaps in the oral presentation rubric by reducing ambiguity and minimizing the subjective elements in the descriptors to aid easy and accurate grading of the speaking presentation. Therefore, refining the speaking presentation rubric should go hand in hand with user training to achieve the desired result. The refinement process should focus on amending the rubric to clarify criteria and standardize expectations, thereby reducing ambiguity for assessors. This may

involve rephrasing descriptors, providing clearer indicators, or including specific requirements such as target lexis and grammatical elements in the descriptors. Additionally, periodic user training, including detailed explanations of the revised criteria, practice assessments, and discussions on mitigating subjective biases during evaluation, can facilitate raters further. This comprehensive approach is expected to improve the reliability and validity of the assessment of oral presentations. Ultimately, this study aims to enhance the fairness and accuracy of grading while maintaining the integrity of the assessment process. In a nutshell, to promote a fair and open evaluation procedure and strengthen the validity of assessments, measures should be taken to yield a more objective scoring across a range of domains and eliminate personal biases.

The existing research provides valuable insights into assessing oral presentations in higher education. However, there are still knowledge gaps that need to be addressed. One potential area for future research is the correlation between perception and the application of the rubric by different examiners and its effect on the scores. Moreover, there is a good opportunity to explore the specific components of rubric-based assessment that affect the raters' perception. Additionally, future research can explore reasons for having a personal impression if the rubrics are administered in the same environment.

REFERENCES

- Ahmadi, A. (2020). Rater dominance in discussion as a resolution method. *Taiwan Journal of TESOL*, 17(1), 141-165.
- Almohaimede, A. (2021). Comparison between students' self-evaluation and faculty members' evaluation in a clinical endodontic course at King Saud University. *European Journal of Dental Education*, 26(3), 569-576.
- Aly, M. M. (2020). Rubrics for EFL oral presentations: A position paper. *Research in Language Teaching*, 10(10), 141-168.
- Belboukhaddaoui, I., & Ginkel, S. (2019). Fostering oral presentation skills by the timing of feedback: An exploratory study in virtual reality. *Research on Education and Media*, 11(1), 25-31.
- Christensen, D., Barnes, J., Rees, D., & Calvasina, G. (2011). Improving the oral presentation skills of accounting students: An experiment. *Journal of College Teaching & Learning*, 2(1), 17-26.
- Creswell, J., & Plano Clark, V. (2011). *Designing and conducting mixed methods research*. Sage.

- Davis, L. (2012). *Rater expertise in a second language speaking assessment: The influence of training and experience* [Unpublished doctoral dissertation]. University of Hawai'i at Manoa.
- Duijm, K., Schoonen, R., & Hulstijn, J. H. (2018). Professional and non-professional raters' responsiveness to fluency and accuracy in L2 speech: An experimental approach. *Language Testing*, 35(4), 501-527.
- Gunderson, J., MacDonald, L., & Gunderson, W. (2021). Integrating rubric-based metacognitive reflection to improve scientific research presentations. *Journal of Chemical Education*, 98(7), 2272-2278.
- Haji Suhaili, W. S., & Haywood, J. (2017). A student-centered approach to ideas generation for projects: Is it a threat to creativity and innovation? *Journal of Nusantara Studies*, 2(1), 13-26.
- Hidri, S. (2018). Assessing spoken language ability: A many-facet Rasch analysis. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 23-48). Springer.
- Huang, Y., & Gui, M. (2014). Articulating teachers' expectations afore: Impact of rubrics on Chinese EFL learners' self-assessment and speaking ability. *Journal of Education and Training Studies*, 3(3), 126-132.
- Mehmet, Ş. A. T. A., & Karakaya, I. (2022). Investigating the impact of rater training on rater errors in the process of assessing writing skill. *International Journal of Assessment Tools in Education*, 9(2), 492-514.
- Murillo-Zamorano, L. R., & Montanero, M. (2018). Oral presentations in higher education: A comparison of the impact of peer and teacher feedback. *Assessment & Evaluation in Higher Education*, 43, 138-150.
- Namin, A., Ketron, S., Kaltcheva, V. D., & Winsor, R. D. (2021). Improving student presentation skills using asynchronous video-based projects. *Journal of Management Education*, 45(6), 987-1010.
- Pérez-Torregrosa, A., Gallego-Arrufat, M., & Cebrián-de-la-Serna, M. (2022). Digital rubric-based assessment of oral presentation competence with technological resources for preservice teachers. *Estudios Sobre Educación*, 43, 177-198.
- Polat, M. (2020). A Rasch analysis of rater behaviour in speaking assessment. *International Online Journal of Education and Teaching*, 7(3), 1126- 1141.
- Reddy, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.

- Thala, T., & Sheehan, S. (2022). The processes of rating L2 speaking performance using an analytic rating scale: A qualitative exploration. *Language Education & Assessment*, 5(1), 34-51.
- Ulker, V. (2017). The design and use of speaking assessment rubrics. *Journal of Education and Practice*, 8(32), 135-141.