Asian Journal of Assessment in Teaching and Learning
*Vol 14, Issue 1, 2024 (117-129) /eISSN **2821-2916***

Using Response Tokens as a Criterion in an Oral Assessment

Rubini Radza[1]*, Afza Diyana Abdullah[1], Qiu XiaoTing[1]
[1]School of Educational Studies, Universiti Sains Malaysia, Pulau Pinang, Malaysia
*Corresponding author: rubyradzaaiulc@gmail.com

Published: 30 June 2024

To cite this article (APA): Radza, R., Abdullah , A. D., & Xiao Ting, Q. (2024). Using Response Tokens as a Criterion in an Oral Assessment. *Asian Journal of Assessment in Teaching and Learning*, *14*(1), 117–129. https://doi.org/10.37134/ajatel.vol14.1.10.2024

To link to this article: https://doi.org/10.37134/ajatel.vol14.1.10.2024

Abstract

This research aims to investigate the use of response tokens as a criterion in an oral assessment. A few research instrument is used in this study to find out the types of response tokens used by English speakers and the reasons behind using them. Besides, this study also helps to determine whether the use of response token has effects on oral assessment and to discover the validity and reliability of their usage. Osgood-Schramm's Model of Communication is used in the assessment; as in to evaluate speaking skills, listening skills and also the use of response tokens as a part of active listening strategies. Listening skills are often assessed on its own without the integration with other skills in most standard testing, thus this research studies the role of active listening in the oral test. An assessment based on interactive speaking and listening model involving the use of response token is carried out and the results are thematically categorised. The findings from this research proved that the response tokens are actively used as a listening strategy for several different reasons or to convey different meanings and the use of response token as a component in assessment are both valid and reliable. The most significant implication of this finding is on assessment and pedagogical aspect where the importance of response token as one of the measuring components in communicative competence when conducting oral assessments is highlighted.

Keywords: *Response Tokens, Oral Assessment, Interactive Model, Speaking Skills, Listening Skills, Active Listening*

INTRODUCTION

Among all four language skills a learner needs to master, listening skill is seldom being given sufficient distinction in many aspects. Several factors could have contributed to this which include limited research studies on listening as well as how listening proficiency is assessed. As a receptive language skill, listening proficiency is often assessed through comprehension level. Even though this is a common practice, it does not represent listening skill as a whole. Interactive listening whereby an individual takes part in a conversation as an active listener is rarely addressed. ESL (English as the Second Language) listening assessment is oftentimes modelled over a linear one-way communication where learners are at the receiving end (Ryan, 2022). High-level English language proficiency tests such as the International English Language Testing System (IELTS), The Malaysian University English Test (MUET) and Test of English as a Foreign Language (TOEFL) conduct the listening test individually whereby listeners are expected to be quiet in order to listen to the audio. In such a setting, they are required to listen attentively to a formatted aural text provided in the form of a recorded native-speaker voice and subsequently, answer a set of comprehension tasks to measure how well they are able to understand the aural text. The underlying reason for this structure is mostly to assess learners' mastery of a set of sub-skills in listening skill for example listening for gist and listening for specific information. The skills are considered essential as they lay the foundation for listening strategies. However, the purpose of listening is beyond getting the information from an aural text. It also includes listening for

interaction purposes. In an interaction, both speakers and listeners carry a part to maintain the flow of the dyadic interaction (Galaczi, 2014; Galaczi, 2008). While speakers are talking, they may need a sign of acknowledgement, affirmation or even disagreement from the listeners to indicate that the content of their message comes through. These signals from the listener are termed as response tokens. Response tokens may be observed in verbal and non-verbal forms (Eiswirth, 2022; Norrick, 2012). When the listener responds with any of the response tokens, he is actively listening to the speaker (Goh & Vandergrift, 2021). This is reflected in most interactive models (Ryan, 2022). Ryan further emphasises that listening assessment should no longer be based on one-way communication instead, the idea of listening in interaction should be applied in the listening assessment. In view of this gap in research studies, this research aims to investigate the use of response tokens as a criterion in an oral assessment.

From the perspective of language testing, the usage of response tokens like "uh-huh," "okay", "yeah", "like", "alright", "oh", and "mm-hmm," has been catching the interest of many, especially in EFL students' context. These tokens may carry ambiguous meaning as they are frequently referred to as linguistic properties which are "unclassified", "homogeneous" and "messy, nevertheless convey more meanings in speaking (Hug & Amir, 2015). The habit of using response tokens as such in speaking tests may influence the learners' scores for fluency and accuracy. The usage of response tokens has the tendency to affect fluency where it disrupts the flow of speech (pauses and hesitations while thinking) and has influence on accuracy as well when the students use them wrongly or in unsuitable context which indicate incertitude. Both of these elements, although seen as two different prominent features in an oral assessment rubric, they are interconnected. The evaluation of speaking skills is quite rigid and normally limited to directed responses, questions about something in particular subject, topics or pictures and reading aloud, which may not be able to reflect on the students' real capacity or proficiency in speaking (Ugiljon, 2018). According to Amar et.al., (2021), previous studies on this matter has proven that students may show certain level of resistance by choosing to provide minimal responses even when they are given oral tests in a clear structured format which means their minimal responses may possibly be used meaningfully rather than just fillers, for instance, using "yeah" to show agreement to something, "right" to shift to another topic or "oh" to show understanding(Ebshiana, 2020). These words can stand on their own and function as short responses to indicate turn-taking in a conversation (Gardner, 2007). This has put the validity and reliability of the assessment techniques at a questionable position to the researchers. The data obtained from a preliminary study conducted involving multinational EFL students in the Language Centre in Albukhary International University demonstrated a prominent use of response tokens in the practice of IELTS speaking test. These students, despite having different proficiency levels in English, used various types of response tokens during the speaking test. The use of response tokens (sometimes overuse) by students may affect their speaking scores to a certain extend but the concern is that, the speaking band descriptors or rubrics (IDP) in CEFR standards generally do not include response token as one of the criteria which has led to this study. Having confirmed its validity and worth, the researchers decided to investigate the use of response token as a criterion in oral assessment as it was considered underexplored due to its lack of previous literature. This study aims to explore further on the types of response tokens which were commonly used, the contributing factors behind this habit, their functions and impacts on speaking skill. In order to clarify the aforementioned aspects, the researchers carried out this research intending to provide valuable insights and make necessary changes in the speaking assessment rubric and pedagogy in the future. Knowing the features, placement and mannerism might lay out their corelation in conversations, to modify the role of teacher and the contribution of the learner (Walsh, 2011a).

LITERATURE REVIEW

1. Response Token

The use of response tokens in interactive communication has been documented in a number of research studies in the literature. Gardner (2001) found brief utterances of response tokens like "uh-huh," "oh," "mm," "yeah," and "right," which may not have traditional meanings but are used by listeners to build sentences. Norrick (2012) uses Conversation Analysis to provide a thorough analysis of the work of linguistic response tokens. According to the research, emotional response cues like "wow" imply a

likelihood of participation in a conversation, while "wow" and "gosh" are cues that indicate more intense reactions to a current encounter. Furthermore, words like "yeah," "oh," and "really" communicate information and evoke a response with various frequencies. The analysis, however, is limited to language tokens by unaware of nonverbal cues like "nodding," "smiling," and "staring," which can be used to measure listening in interactive communication, listeners can refrain from interrupting speakers in an effort to be courteous and encourage them to continue speaking. Some authors examined how the various intonations used in answer tokens could influence an interactive dialogue. For instance, Sørensen(2021) did research in a Danish setting, and analysis of the data showed that tokens with a rising tone were used when listeners tried to validate the information. Additionally, a rising tone developed when confirmation, appraisal, and approval were added. Due to limitations in the Danish context rather than the English environment, the research concepts may lack applicability, but it expands the research by a significant response to the intonation regions. Some surveys have concentrated just on a single typical response token. De Stefani (2021) looked at instances where head nods were produced in advance as reflected and latent responses to polar questions. The results of this study revealed that listeners can express their rights by silently nodding when a question is being asked. Additionally, it discovered that nodding in conjunction with a verbal expansion may help listeners go from interjection to a transformative response. The French and Italian video corpus, which was taken from typical and institutional multiparty encounters, may be the study's weakness because they have less instructional implications than English. While most studies prefer to elaborate on the positive response tokens in contrast to a few evaluations on the negative response tokens, communication covers numerous settings such as daily life, study, and business. They targeted business speech, which focuses on the negative reaction token “no” in English and “net” in Russian, as Malyuga & McCarthy (2021) have noted. Two corpora of spoken business or professional conversation were used to collect the research data in order to build functional comparability and distinguish them in both contexts (English and Russian). The results from the authors' use of corpus linguistics, conversation analysis, and discourse analytical techniques show that “no/net” performed a wide range of functions related to conversational continuity, subject management, turn-taking, and hedging.

2. Interactive Model

When an individual speaks, he builds a connection to communicate. The speaker becomes the encoder to the message he conveyed and the listener therefore, is the decoder of the message. In this linear understanding of communication, all three components are fundamental for communication to occur: encoder, message, decoder (Bhatnagar, 2011). The encoder usually encodes a message with a purpose (Paredes, 2021; Encalada & Sarmiento, 2019). Any type of communication is purposive as it may carry transactional, informational, social, or entertainment purposes to the listener. In contrast, a communication that does not serve purposeful meaning to the receiver can be deemed slightly superficial (Ryan, 2022; Ulker, 2017). In effect, this superficiality feature does not support language learners in situations that demand real communication. Generally, many types of oral assessment (MUET, IELTS, SPM – Speaking test) carry this superficial feature whereby the test-taker's purpose is to gain high marks, resulting in an interaction that appears shallow. In a speaking test that includes assessing the interaction of at least two candidates, the dyadic interaction should be assessed based on the speaker-listener dimension, not bound to the speaking skills and communicative competence of the speaker. This is because both speaker and listener direct the flow of exchanges. Inherently, the listener’s role in the interaction should also be considered as it contributes to the continuity of the communication (Ryan, 2022). A communicative approach to oral assessment is more appropriately analysed through the use of an interactive model. There are various interactive models of communication: Laswell’s Model, Shannon and Weaver’s Model (1949), Bolton and Cleaver’s Model (1949), Wendell Johnson’s Model (1951) and different versions of Schramm’s Model. Each model takes on a different perspective and interpretation of communication. However, a communicative approach of oral assessment requires a particular shape of interactive model. In this case, a circular model is more appropriate as it shows that a message does not only end when it reaches the receiver. Instead, it becomes the building blocks to the continuity of the conversation/discussion. Therefore, Schramm’s (1954) Interactive Model is a suitable model to view oral assessment through a communicative approach. The difference between Osgood-Schramm’s 1954 model and the helix version of Osgood-Schramm model (Bhatnagar, 2011,

p. 97) is the elements of Personality and Noise. Removing these two may not influence an oral test that has limited time allocation.

Figure 1 below shows the Osgood-Schram's Model of communication. As illustrated in the diagram, in the first circle, a speaker (who is also the listener) carries out the role as an encoder, interpreter and decoder of the message. In other words, when the speaker conveys his message, he becomes the encoder. This message does not reach straight to the intended listener. Instead, it is presented in a display of verbal production of language, facial expressions and body language. The intended listener will de-code this presentation of the message and interpret it. Subsequently, the listener becomes the encoder and encodes the message to the first speaker who will now take up the role as a listener and decode said message. The listener in this model does not passively interpret the message, but he actively interprets the message displayed. Hence, both speaker and listener carry the responsibilities equally in a communication between two people. In this research, the response to an encoded message can simply be in the form of a response token. Norrick explains that "*in producing minimal response tokens, a listener signals a willingness to remain (predominantly) silent, to refrain from interrupting and to attend to the primary speaker, and thereby encourages the speaker to continue with a multi-unit turn*" (2012, p. 575).

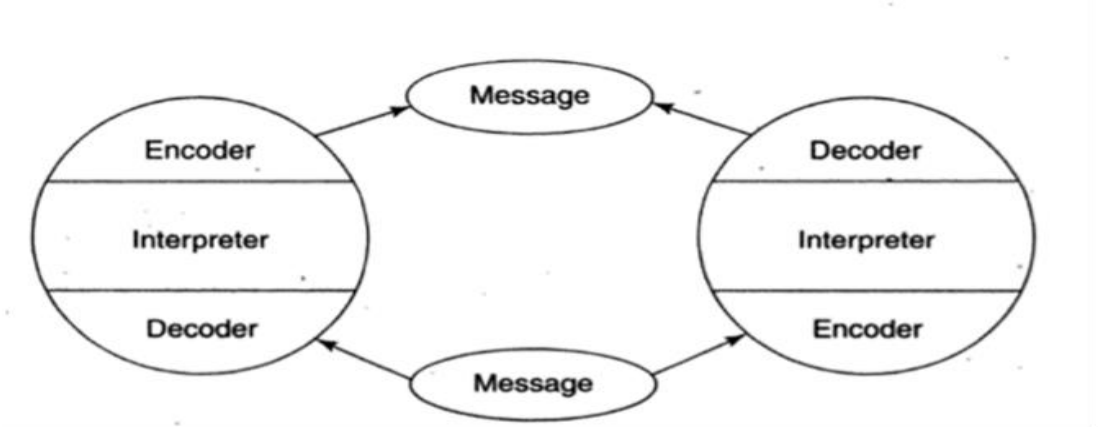

Figure 1 Osgood-Schramm's Model of Communication from 1975 to 2000

From the assessment perspective, evaluating speaking skill and listening skill (including response tokens) in a dyadic interaction can be a complex undertaking by only one assessor. While some (CEFR), the dyadic interaction evaluation can be found in other CEFR-based tests such as Cambridge ESOL (English for Speakers of Other Languages) (University of Cambridge, 2009). Therefore, its usability for adapting to the active listener's dimension (in this case exchanges can be pre-planned, there are still instances of spontaneity that may appear to be difficult to capture. Therefore, it is essential to establish the assessment requirements which would allow a more comprehensive evaluation of a discussion. Other than that, it should also include an assessment that prioritises both speaking and listening components of the oral test. In this research, the listening component will include identification of response tokens strategies that the listener uses.

3. Assessment

The general rule of a good assessment is it should be valid and reliable. In one aspect, the validity of an assessment explores the extent of which an assessment assesses what it intends to assess. In terms of creating assessment rubric, Moskal & Leydens (2000) argues that, *"validation is the process of accumulating evidence that supports the appropriateness of the inferences that are made of student responses for [specific] assessment uses*" (p. 1). On the other hand, the reliability of an assessment refers to the ability of the assessment to reproduce consistent results. In view of this rule, large-scale speaking assessments strive to achieve a good level of validity and reliability such as TOEFL, IELTS, MUET and SPM. However, these examples of exams do not explicitly address the impact of response tokens as part of an active listener response to a message in a communication model. In TOEFL, IELTS

and MUET, the candidate is individually evaluated on his speaking abilities whereas the SPM Speaking test addresses the interaction model in the communicative competence dimension. Following the Common European Framework of Reference inclusion of response token strategies) would be apt. Most oral assessments use rubrics to evaluate different dimensions of interaction in the target language. A rubric assessment is defined as criteria for measuring performance dimensions or standards of achievement (Panadero & Jonsson, 2020; Jonsson & Svingby, 2007). In the SPM Speaking Assessment rubric, the evaluation is carried out by two people: the assessor and the interlocutor. The assessor will carry out analytical assessment of grammar, vocabulary and communicative competence, while the interlocutor will assess the overall performance. This assessment is comprehensive in this format because it addresses the dimensions via holistic and analytic rubrics (Panadero & Jonsson, 2020). Focusing on the interactive communication element in the Communicative Competence dimension, it highlights how a speaker initiates, maintains and closes an interaction. In addition, it also refers to a candidate's ability to maintain a natural-flow of a conversation despite hesitation to make meaning clear and clarify the meaning of others (Lembaga Peperiksaan Malaysia, 2021). Adapting to this rubric, the research proposes the inclusion of the interactive listening dimension with response token strategies as one of the criteria in the analytic rubric of SPM Speaking Test. An expansion of constructs in this dimension will include: Listener behaviour, active role of listener, communicative purpose, and collaborative nature (Xiaoxian & Yan, 2010). This is reflected in the dimension through the use of response strategies and attention to the speaker.

METHODOLOGY AND DATA COLLECTION

1. Participants

The participants were 42 General English course students from Albukhary International University, Language Centre, Malaysia. There were 26 males and 16 females, ages 19 to 23. The participants came from different countries and had a mixed level of proficiency in English language, from beginner to intermediate level. Some of them had studied English in primary and high school, while others had only recently started learning English prior to the assessment. The General English course was skill-based (Reading and Writing, Listening and Speaking), which logically synchronised with the design of this research.

2. Research design

The research is conducted to investigate the validity and reliability of using response tokens as a criterion in an oral assessment. In order to address the first research question, a mixed-method approach was used. The study consisted of a pilot study, where two groups of international students with 3 students in each group were asked to conduct a spontaneous discussion on a topic without any preparation beforehand. The conversations were video-recorded and thoroughly transcribed and assessed by the researchers to find out the types of response tokens used by L2 speakers. This was followed by discussions in pairs about two (2) different general speaking topics extracted from IELTS Part 1 Speaking themes. Subsequently, interviews were conducted to discover the types of response token used by L2 speakers, the reasons why L2 speakers use them and to ascertain whether the use of response token is valid and reliable in oral assessment. Selection of candidates for the interviews was done based on their proficiency levels according to the scores given after assessing their conversation videos. Data collection was done through qualitative and quantitative instruments (eg. group based discussions, discussions in pairs and one-to-one interviews with a teacher).

3. Instrument

a. Conversation/Discussion (Video Recording)

The conversations were based on IELTS Part 1 Speaking themes, each pair was given 2 topics to discuss on. The participants were asked to take a video recording (not exceeding 10 minutes) of their conversations and submit the recorded videos on Telegram to the teacher personally. For each theme or

topic, there were approximately 5-7 questions and students were allowed to modify, omit or combine some questions depending on the necessity and the flow of their dialogue. Twenty-one (21) videos were chosen for this study and were evaluated using the SPM Oral Assessment rubric.

b. Interview (Audio Recording)

This was a one-to-one interview with the teacher where the interviewees were asked a few follow-up questions related to active listening and the use of response tokens. There were a total of eight (8) questions asked by the teacher to determine the reasons behind the use of response tokens in accordance with research question 2. The structured questions were prepared by the research team after a thorough discussion. For the interview sessions, two students from different levels were identified and put into 3 different categories (beginner, pre-intermediate and intermediate). For every category, one male and one female student were picked out respectively which summed up to 6 students. The interviews were scheduled separately with each one of them and the teacher estimated about 10 minutes maximum to be spent on one interviewee. The interviews were audio-recorded by both the interviewer and interviewee.

c. Assessment Rubric

In this research, an oral assessment rubric adapted from Sijil Pelajaran Malaysia (SPM) Speaking Test. It is an analytic rubric. Originally, the rubric consisted of three (3) components namely Grammar, Vocabulary and Communicative Competence (Speaker). An extra component which was the Communicative Competence (Listener) was added onto the rubric to best suit the requirements of this study. As the research revolved around both Listening and Speaking skills, the research team was positive that it was important for the extra component to be added. The band score for this assessment rubric ranged between zero (0) being the lowest and six (6), the highest for each component. The maximum overall score added up for all components would be 24, and the minimum, 0.

In order to prove the construct validity of an oral exam, the research hypothesised that a variety of features, or subconstructs, such as grammar, vocabulary and communicative competence, are included in the interactive oral ability (or construct) itself. Then, using an oral test with those traits, a pilot test is conducted on a group of students (Reddy & Andrade, 2010). The scores should be reliable before the internal correlations of the subconstructs, the relationships between each subconstruct, and the relationships between the subconstruct and the construct can be determined. The oral test is stated to have high construct validity when the coefficients between distinct speaking ability traits are low but the coefficients between the overall construct and each subconstruct are high. Reliability of a rubric can be categorised into two: inter-rater reliability and intra-rater reliability (Reddy & Andrade, 2010; Moskal, 2000). Inter-rater reliability refers to the possibility that students' marks may differ from one rater to another, while intra-rater reliability refers to the changes in the rater scoring due to external factors. In this research, Inter-rater reliability is achieved by giving briefing to all the three raters. Later, scores were verified between the three raters. On the other hand, the intra-rater reliability is controlled by giving 5 days to complete 12 students' oral assessment.

d. Listening strategy checklist

A checklist of listening strategies is constructed based on Norrick's (2012) list of listening strategies that listeners use in an interactive communication. The listening strategies listed are based on functions of response tokens that have been based on various studies on response tokens. This checklist is used to identify the types of response tokens used by the sample in the video recording. The videos are analysed and the strategy used by the listener is identified based on frequency. If the listener uses a strategy more than 3 times, "F" (Frequent) is labelled, once or twice, "S" (Seldom) and "N" (None) when the strategy is not used.

4. Procedures

As the utmost initial step of the research, a pilot study was conducted to find out the types of response tokens used by L2 speakers. A *pilot study* asks whether something can be done, should the researchers proceed with it, and if so, how. However, a pilot study also has a specific design feature; it is conducted on a smaller scale than the main or full-scale study (In, 2017). Six (6) international students who were in intermediate level of English proficiency were hand-picked by the teacher to get involved in the pilot study. These students were undergraduate students from School of Business in Albukhary International University who had already passed their IELTS test with a minimum of 5.5 band score and had a considerable level of fluency. The students were assembled in 2 different groups (3 students in each group). The teacher described the nature of the task and instructed the students to have a casual and spontaneous conversation about a topic (The topic given was future plan). Their conversations were video-recorded, and the topic of the discussion was only given once the recording had begun. At this point, the teacher leaves the group on their own to avoid distractions. This was done to prevent the *Hawthorne effect* which refers to the tendency for study participants to change their behaviour simply as a result of being observed (James & Vo, 2010). Once they're done, the recordings were shared on Telegram with the teacher, and the teacher shared the recording with the research group members. The recorded discussions were transcribed to check the use of verbal response tokens and watched to check the non-verbal response tokens. The pilot study was considered a success since the researchers were able to detect the types of response tokens used by L2 speakers that directly answered research question 1 and proved the validity of the research topic. This study was carried out in a natural classroom setting. Firstly, students from four (4) different classes were asked to work in pairs preferably with someone from a different class and/or another country. Once they are seated in pairs, the teacher explained the activity clearly on what is expected from the students and opened room for clarifications and provided necessary guidance for the students before leaving them with their pair to complete the speaking task. Before the activity started, the students were asked to fill in some details about themselves and sign the consent form in order to follow a general protocol as some of their recordings would be used for the research. After the collection of consent forms, the teacher handed each pair of students 2 sets of IELTS questions related to the discussion topics for their reference. All pairs had different topics and questions. This was mainly done to assist the weaker students in their conversations yet, they were given the privilege to modify the questions if they found it necessary. The students then had the freedom to choose the venue (study hall, library, classroom etc.) to have and record their discussions as it would not be ideal to record a speaking activity inside a crowded classroom. In pairs, they were required to record a video (less than 10 minutes) and share the videos on Telegram personally to the teacher. Only one student had to submit the video with the details (such as name and class) of the other partner. Twenty-one (21) videos were divided among three researchers (seven [7] each) and assessed based on SPM Oral Assessment Rubric. The score of the interviewees were tabulated for further discussion and analysis. Proceeding from the conversation videos, the third stage of this study was carried out. It was pre-planned with structured questions about active listening and the use of response tokens. The teacher allocated roughly 10 minutes per interview. The interviewees were named based on their performances on second stage (conversation video). Two students (male and female) were chosen from three different levels (beginner, pre-intermediate and intermediate). They were personally informed by the same teacher about the interview session, followed the schedules and completed the interviews. The interviews were audio-recorded, by both the interviewee and interviewer. After the interview, the interviewees were asked to submit the recorded audio file to the teacher on Telegram. The six (6) interviews were divided among the researchers in the ratio of two to one (2:1). The audios were transcribed by the researchers for further thematic analysis. The themes were generated based on the structured questions and the general pattern of the responses gained from the interviews. Purposive sampling method was used throughout the study given the uniqueness of this research area. The research team came to an agreement that it would best to use purposive sampling as it would be sufficient to cover the research objectives and to answer the research questions. The reason for *purposive sampling* is the better matching of the sample to the aims and objectives of the research, thus improving the rigour of the study and trustworthiness of the data and results. Four aspects to this concept have previously been described: credibility, transferability, dependability and confirmability (Campbell et. al., 2020).

RESULTS

1. Frequency of Listening strategy

Table 1 shows the frequency of listening strategies used by the sample. Overall, the highest frequency of listening strategy used is Maintenance/Sustaining/Feedback (F = 41). All the participants used this strategy to maintain the conversation. Some of the examples of response tokens used in this strategy are *"Yeah"*, *"Yeah"*, *"Uh-huh"*, *"Mm-hmm"* [Nod] and [Smile]. This is followed by Maintenance/Sustaining/Reactors (F=28) and Maintenance/Sustaining/Follow-ups (F=24). For the former, responses include *"Wow"*, *"Good"*, *"That's interesting"* and the latter display responses such as [A: *"I like tea."*, B: *"Why tea?"*] and [C: *"Volleyball."*, D: *"When did you first play?"*]. These three strategies are frequently used and it shows that the participants are able to show participation in the conversation. Next, the participants use Maintenance/Sustaining/Shadowing (F=16) and Interpretive Summary (F= 11). For Maintenance/Sustaining/Shadowing, the participants repeated some of the words used by the speaker as soon as it was uttered. On the other hand, Interpretive Summary shows a higher form of active listening as the listener summed up what he understood from the conversation before it carries on. The first four strategies are used frequently indicating that they are lower level listening strategies. As the listening strategies do not require much interpretation from the message encoded by the speaker. There are also some other factors that should be taken into account. One of which is that the conversations were not natural and mostly are practised beforehand. Hence, the strategies are used to acknowledge that message is received in more of assessment of the message and not interpretation of the message. This is also the reason why the first three strategies which request for clarity of unclear words/phrases are not used in this setting (except for 1 time by a participant who could not hear the speaker's encoded message due to volume). In a dyadic interaction of Osgood-Schramm's communication model, even a premeditated one, it is important for the listener to decode the message from the speaker and encode the message to show acknowledgement that the message has come through. This cyclical process encourages the conversation to carry on.

Table 1 Frequency of Response Token Strategy

No	Response tokens strategy	Frequency		
		None	Seldom	Frequent
01	Request provide repetition	41	1	0
02	Expressing non-understanding	42	0	0
03	Expressing unknowingness	42	0	0
04	Request provide clarification	30	0	0
05	Interpretive summary	17	14	11
06	Maintenance/Sustaining/Reactors	5	9	28
07	Maintenance/Sustaining/Feedback	0	1	41
08	Maintenance/Sustaining/Shadowing	16	10	16
09	Maintenance/Sustaining/Follow-ups (to an answer, develop with WH question)	16	10	16
10	Avoidance behaviour	39	3	0

Two other listening strategies skills which were not used quite as frequent are Request Provide Repetition and Avoidance Behaviour. The former is used once by a participant due to inaudible words from the speaker. On the other hand, the Avoidance Behaviour shows that the participants are not actively involved in the conversation. The use of this listening skill shows that the listener is a passive listener. In the Osgood-Schramm's Communication Model, the listeners did not decode the speakers' message but still sent a message to the speaker pretending that they understood. Norrick (2012) warns that the use of this strategy may show that the listener does not understand the topic of the conversation but wants to appear agreeable to the speaker. Prolong use of this strategy will result in communication breakdown. The pedagogical implication from this is to teach learners how to navigate a conversation as an active listener using other listening strategies apart from the Avoidance Behaviour.

2. Validity of Oral Assessment

Table 2 shows the correlation between the constructs in the oral assessment. Based on the table, the result shows that Communicative Competence as Listener construct correlates positively with other constructs: Grammar (r = .628, p = < 0.001), Vocabulary (r = .574, p = <0.001) and Communicative Competence as Speaker (r = 651, p = <0.001). The correlation between Communicative Competence as Listener and the other 3 constructs are significant. Hence, the validity of this oral assessment using the newly added construct (Communicative Competence as Listener) is established.

Table 2 Correlation Between the Constructs in Oral Assessment

		Grammar	**Vocabulary**	**Communicative Competence: Speaker**	**Communicative Competence: Listener**
Grammar	Pearson Correlation	1	.623**	.496**	.628**
	Sig. (2-tailed)		< .001	< .001	< .001
	N	42	42	42	42
Vocabulary	Pearson Correlation	.623**	1	.465**	.574**
	Sig. (2-tailed)	< .001		.002	< . 001
	N	42	42	42	42
Communicative Competence: Speaker	Pearson Correlation	.496**	.465**	1	.651**
	Sig. (2-tailed)	< . 001	.002		< .001
	N	42	42	42	42
Communicative Competence: Listener	Pearson Correlation	.628**	.574**	.651**	1
	Sig. (2-tailed)	< . 001	< .001	< .001	
	N	42	42	42	42

** Correlation is significant at the 0.01 level (2-tailed).

3. Reliability of Oral Assessment

Table 3 shows the reliability of the Oral Assessment using Cronbach's Alpha test of reliability. The analysis is carried out on four items: Grammar, Vocabulary, Communicative Competence: Speaker, Communicative Competence: Listener. According to the table, the Oral Assessment achieved the Cronbrach's Alpha value of .843 which indicates high internal reliability. This further implies that the test control of inter-rater reliability is successfully achieved. Hence, the oral assessment rubric can be used when conducting the assessment which includes Communicative Competence of the Listener.

Table 3 Reliability of Oral Assessment

Cronbach's Alpha	**Cronbach's Alpha Based on Standardized Items**	**N of Items**
.843	.843	4

4. Type of listeners

All of the respondents to the interview stated that they can participate actively in a conversation with friends, demonstrating that they are active listeners. The participants can easily alternate between speaking and listening in a casual setting. Both responsibilities create an interactive process and build effective communication.

5. Listening strategies

The students can contribute to the communication verbally or by body language (such as eye contact, nodding, and facial expressions). For example, S1 said: "*Hmm, I think that facial movement is really important, but I try not to overdo it. I nod sometimes when they speak and I try to maintain eye contact, and also I feel like it's important to listen, not to respond back but to understand what they're saying."* According to Osgood-Schramm's Model of Communication (1954), the speaker encode the message to interpret it, and then the listener will decode the message and in turn, the listener will encode a message by using verbal or non-verbal response tokens to show their attention, acknowledgement, agreement or surprise which will encourage the speaker to continue the communication.

6. Perceived use of Body Language

The respondents stated that they would communicate with people through making eye contact, nodding, and using hand gestures. Some of them would use it frequently, while others insisted that it should only be used sparingly (not too much or too little). S2 gave his example: "*I used to nod a lot. Whatever they say I used to nod, even if they are wrong I used to, I try to [like] motivate them, I used to use yeah, you're right."* Nodding as one type of nonverbal response tokens was studied in the previous studies, for instance, Stefani (2021) concluded that the listeners can express their rights by silently nodding. Nodding implies someone is a good listener.

7. Response Token Situations

Verbal or non-verbal response tokens will be used in different situations for the respondents. When asked by the speakers, they will use response tokens to maintain conversation. Response tokens have occasionally been used to satisfy the expectations of speakers, as well as to demonstrate recognition and relevancy. Also, the response tokens will be used when the listeners want to display their agreements, according to the interviewees. For example, S3 stated: "*When I do it to emphasise, like if they want me to agree or if I agree with something I do it, if it's something I could relate to I do it, yea I think so, that's what I do most.".* All of these situations were brought up by the participants, and they are consistent with the findings of Norrick's (2012) study, which suggested that when taking part in a conversation, certain response tokens may indicate a reaction to the speaker and prompt the speaker to provide more information.

8. Implications for not responding: Cultural Influence

The majority of respondents agree that it is improper if one does not respond to the speaker regardless of their cultural influences. They think that it can be rude and there are possibilities to make the speaker angry or unhappy to continue the conversation. While, some strongly concurred that not responding to the speaker is culturally unacceptable. For example, S5 mentioned: *"They might be bored, end the conversation, it's culturally rude and displays attitude problems.".* When being asked the same question, S6 said: *"It's not nice, disrespectful in every culture in my point of view."*. Based on these responses it is comprehensive that there is a coherent connection between cultural influence and interactions. Therefore, respondents were aware of the importance of responding and their actions were clearly affected by their cultures. This can be validated by this statement: Research has shown that national culture influences an individual's perceptions, behaviour and beliefs (Harrison & Huntington, 2000; Hofstede, 2001; Kirkman et. al., 2006).

9. Importance of Eye Contact

Eye contact is an essential non-verbal feature in any conversation and it is often perceived as a way of showing attention and respect when interacting with an individual or more people. As per Kleinke's argument, eye contact is a fundamental aspect of nonverbal communication and social interaction from birth throughout adulthood (Kleinke, 1986). In regard to this notion, the interviewees were asked whether it is necessary to maintain eye contact in a conversation, all respondents in general acknowledge the vitality of its role to a certain extent in showing understanding, connection,

participation, respect and focus. For instance, S1 said: *"..eye contact is very important when you look at the speaker they will know you are listening.."*, while S3 said: *"..of course because eye contact is a way to understand the person with conversation.. yeah to show focus.."*. In addition to this, two more students suggested that it is important not to overdo it when it comes to eye contact. They believed that it is good to use eye contact but not all the time. For example, S5 stated: *"..same gender is like ok but opposite gender when you try to speak, too much staring, they'll be like ... even if you speak casually, they'll be uncomfortable, we too feel awkward so we need to drop... in the middle we need to take a break to/and avoid."* and S6 said: *"I think eye contact is good, is important but not necessary coz even if you're not looking at them, maybe if they're looking at you they'll know that you nod your head even if you not look at them, and when you give that nod they'll know you're listening."*. The statements from S5 and S6 seem plausible in a logical context and imply that it's necessary to use eye contact mindfully.

10. Listener's Image

Regarding the perception of the respondents on how they want the speaker to perceive them based on response tokens, all of them indicated that they try to send significant messages to the speaker during a conversation. According to them, any kind of response tokens either verbal or non-verbal are used to signal some sort of understanding or not understanding, agreement or disagreement and showing high or low level of interest in the subject or matter that's being discussed. For example, S5 said: *"If I say yes, yeah yeah that's right, then I want them to take me as like 'oh, he really meant, he really knows about it and he had experience and like okay so he is interested in the topic and he knows a lot about this.."* and to be perceived otherwise *"they can definitely see from my body language that I don't understand what they're trying to say. So, they will try to explain further. I would just say verbally or they can see through my facial expression especially if I frown (chuckles)."*. All respondents gave positive remarks from the speaker's side that whenever they try to use verbal or non-verbal cues, they are well perceived and understood.

11. Other relevant issues

Gender was one issue that was not discussed among the researchers from the preconception of this study until the interviews were conducted. Based on this study it is arguable that boundaries of eye contact when communicating with opposite sex has to be kept in mind to ensure the comfort of the other individual. One of the respondents believed that it may be overwhelming if too much eye contact is used with the opposite gender. Referring to S5's statement: *"..same gender is like ok but opposite gender when you try to speak, too much staring, they'll be like ... even if you speak casually, they'll be uncomfortable, we too feel awkward so we need to drop... in the middle we need to take a break to/and avoid."*. S5 suggested that, if situations like this occur, the individual could try to steer clear of staring or giving eye contact or to take a break to assure that the other party feels better.

CONCLUSION AND RECOMMENDATION

The research is able to answer all the research questions. In terms of types of response tokens used by L2 speakers, it is found that most listeners use the strategy to maintain conversation by giving feedback, reacting and shadowing. Other than that, the result shows that learners use response tokens for various reasons but mainly to show understanding, focus, agreement and connection. Furthermore, the research shows that using response tokens (listening strategies) as part of the oral assessment is proven to be both valid and reliable. The main pedagogical implication from this research is to consider using response tokens as a measurement of listener's communicative competence in an oral assessment. For future studies, research can be carried out focussing on either verbal or non-verbal response tokens and their obtrusiveness.

ACKNOWLEDGEMENT

The authors would like to thank the Director of Albukhary International University Language Centre for permitting the study to be carried out in the language centre teacher's office and classrooms and also for allowing the students to participate in the study.

FUNDING

There was no financial support received for the research, authorship, and publication of this article.

DATA AVAILABILITY

The authors have the permission to share data such as the transcriptions of conversations, discussions and/or interviews upon request without revealing the names of the samples. However, the video and audio recordings have been declared private and confidential.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Amar, C., Nanbu, Z., & Greer, T. (2021). Proffering absurd candidate formulations in the pursuit of progressivity. *Classroom Discourse*, *13*(3), 264–292. https://doi.org/10.1080/19463014.2020.1798259

Bhatnagar, N. (2011). *Effective Communication and Soft Skills*. Pearson Education India.

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of research in nursing : JRN*, *25*(8), 652–661. https://doi.org/10.1177/1744987120927206

De Stefani, E. (2021). Embodied Responses to Questions-in-Progress: Silent Nods as Affirmative Answers. *Discourse Processes*, *58*(4), 353–371. https://doi.org/10.1080/0163853x.2020.1836916

Ebshiana. A., (2020)Response Tokens and Their Sequential Action in the Teacher Third Turn (Note 1) A Conversation Analysis Case Study in the EFL Classroom.

Eiswirth, M. E. (2022). Developing and testing interaction-based coding schemes for the analysis of sociolinguistic variation. *Language & Communication*, *87*, 11-28.

Encalada, M. A. R., & Sarmiento, S. M. A. (2019). Perceptions about Self-recording Videos to Develop EFL Speaking Skills in Two Ecuadorian Universities. *Journal of Language Teaching & Research*, *10*(1).

Galaczi, E. D. (2008). Peer–peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly*, *5*(2), 89-119.

Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests?. *Applied linguistics*, *35*(5), 553-574.

Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. *Language Testing*, *27*(4), 585–602. https://doi.org/10.1177/0265532210364049

Gardner, R. (2001). *When listeners talk : response tokens and listener stance*. Amsterdam Benjamins.

Gardner, R. (2007). The Right connections: Acknowledging epistemic progression in talk. *Language in Society*, 36(3), 319-341.

Goh, C. C., & Vandergrift, L. (2021). *Teaching and learning second language listening: Metacognition in action*. Routledge. In J. (2017). Introduction of a pilot study. *Korean journal of Anesthesiology*, *70*(6), 601–605. https://doi.org/10.4097/kjae.2017.70.6.601

Huq, R.-U., & Amir, A. (2015). When the Tokens Talk: IRF and the Position of Acknowledgement Tokens in Teacher-Student Talk-in-Interaction. Novitas-Royal (Research on Youth and Language), 9(1), 60-76.

IELTS. (2024) IDP IELTS; and Cambridge University Press & Assessment. Retrieved from https://ielts.idp.com/malaysia/results/scores/speaking

James, L., & Vo, H. (Eds.) (2010). Hawthorne effect: Encyclopedia of Research Design . (Vols. 1-0). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781412961288

Jongerius, C., Hessels, R. S., Romijn, J. A., Smets, E. M. A., & Hillen, M. A. (2020). The Measurement of Eye Contact in Human Interactions: A Scoping Review. *Journal of Nonverbal Behavior*, *44*(3), 363–389. https://doi.org/10.1007/s10919-020-00333-3

Jonsson, A. & Svingby, G. (2007), "The use of scoring rubrics: Reliability, validity and educational consequences", Educational Research Review, 2(*2*), pp. 130-144

Li, W. (2011). Validity Considerations in Designing an Oral Test. *Journal of Language Teaching and Research*, *2*(1), 267–269. https://doi.org/10.4304/jltr.2.1.267-269

Malyuga, E. N., & McCarthy, M. (2021). "No" and "net" as response tokens in English and Russian business discourse: In search of a functional equivalence. *Russian Journal of Linguistics*, *25*(2), 391–416. https://doi.org/10.22363/2687-0088-2021-25-2-391-416

Moskal, B., & Leydens, J. (2000). Scoring Rubric Development: Validity and Reliability. *Practical Assessment Research and Evaluation*, *7*.

Neyer, Anne-Katrin & Harzing, Anne-Wil. (2008). The impact of culture on interactions: five lessons learned from the European Commission. European Management Journal. 26. 325-334. 10.1016/j.emj.2008.05.005.

Norrick, N. R. (2012). Listening practices in English conversation: The responses responses elicit. *Journal of Pragmatics*, *44*(5), 566-576.

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*

Reddy, Y. M., & Andrade, H. (July 01, 2010). A review of rubric use in higher education. Assessment & Evaluation in Higher Education, 35, 4, 435-448

Ryan, J. (2022). Listening in interaction: reconceptualizing a core skill. *ELT Journal*.

Sørensen, S. S. (2021). Affiliating in Second Position: Response Tokens with Rising Pitch in Danish. *Research on Language and Social Interaction*, *54*(1), 101–125. https://doi.org/10.1080/08351813.2020.1864159

Ugiljon, A., Anakhon, I., Gulnoza, S. (2018). The Effective Speaking Testing Techniques in Teaching English. *International Journal of Secondary Education*, 6(1), 24-28. https://doi.org/10.11648/j.ijsedu.20180601.15

Ulker, V. (2017). The Design and Use of Speaking Assessment Rubrics. *Journal of Education and Practice*.

University of Cambridge. (2009). *Examples of speaking performance at CEFR levels*. Retrieved from https://www.cambridgeenglish.org/images/22649-rv-examples-of-speaking-performance.pdf

Xiaoxian, G., & Yan, J. (2010). Interactive Listening: Construct Definition and Operationalization in Tests of English as a Foreign Language. *Chinese Journal of Applied Linguistics* (Foreign Language Teaching & Research Press), 33 (*6*)

Walsh, S. (2011a). Exploring classroom discourse: Language in action. Taylor & Francis.https://doi.org/10.3138/cmlr.67.4.536