EXPLORING THE EFFECTIVENESS OF SELF-ASSESSMENT ON MALAYSIAN UPPER SECONDARY ESL STUDENTS' SPEAKING PERFORMANCE

[1]VANESSA JACQUELINE LIGADU
[2*]PRISCILLA SHAK

[1&2] Centre for the Promotion of Knowledge and Language Learning, Universiti Malaysia Sabah,
Jalan UMS, 88400 Kota Kinabalu, Sabah.
Corresponding author: pshak@ums.edu.my

Abstract *Self-assessment plays a vital role in enhancing learning and encourages students to reflect on their learning progress, promotes autonomy in the learning process, and fosters a commitment to mastering the English language. Previous research has highlighted that Malaysian secondary school students face challenges in spoken communication and a review of existing studies suggests that self-assessment can significantly enhance students' learning, specifically in speaking. However, there remains a gap of research on the use of self-assessment in developing speaking skills at the secondary education level. This study seeks to fill this gap by utilising the quasi-experimental, one-group pre-test-post-test design to investigate the effect of the Speaking Self-Assessment Instrument (SSAI) on the speaking performance of 151 Malaysian English as a Second Language (ESL) Form 4 upper secondary students from three schools in Sandakan, Sabah, Malaysia. The students underwent a speaking pre-test based on the Malaysian Certificate of Education (SPM) speaking test format and had their performance rated by two raters to establish inter-rater reliability. The students participated in four sessions of SSAI rubric training, after which they were tested again for a speaking post-test. Data in the form of speaking test scores were compared. The findings showed that the SSAI was most effective for lower-intermediate students, who showed the greatest improvement, while upper-intermediate students showed the least improvement in speaking scores due to their already high proficiency.*

Keywords: *Self-assessment, speaking performance, rubrics, Malaysian secondary students, quasi-experimental*

INTRODUCTION

The Malaysian Ministry of Education (MOE) introduced the Malaysian Common European Framework of Reference for Languages (CEFR) aligned curriculum in year 2017 to help Malaysian ESL secondary school students improve their communication skills and prepare them to be globally competitive (MOE, 2015; Mohd Don & Abdullah, 2019). This initiative has been undertaken to transition from the previous curriculum that focussed heavily on reading and writing skills (MOE, 2015; Mohamad Marzaini et al., 2024) to a more holistic approach in language learning which includes speaking and listening skills (Council of
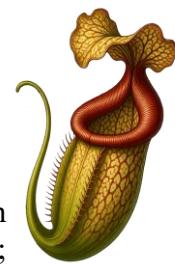
Europe, 2020; MOE, 2015). Pedagogically, the CEFR-aligned curriculum adopts a trinitarian viewpoint whereby teaching, learning and assessment are interconnected (MOE, 2015; Piccardo & North, 2019; Council of Europe, 2020). Accordingly, transparency and continuity in assessment are imperative as they play significant roles in improving students' language skills (Andrade, 2019; Kumar et al., 2023; Moss & Brookhart, 2019).

Prior to the implementation of the curriculum reform in the Malaysian secondary education, the instructional approach emphasises fostering students' skills in anticipation of the SPM examination (Mohamad Marzaini et al., 2024; Supramaniam et al., 2020), a form of summative assessment (Brown, 2001; Hosseini & Nimehchisalem, 2020). While summative assessments offer practicality (Hilden et al., 2022), it conflicts with the focus of the present curriculum on ongoing evaluation of students' progress instead of a single end outcome (Franchis & Mohamad, 2023; Mohd Don & Abdullah, 2019). Frequently, the instructional approach employed leads to repetitive and memorisation-based learning methods, resulting in students who possess little proficiency in the English language, particularly in terms of their oral communication abilities, even after an extended period of 11 years of learning ESL (Nadesan & Md Shah, 2020).

The acquisition of speaking skills has consistently posed a barrier for Malaysian students learning English as a Second Language (ESL) (Kashinathan & Aziz, 2021; Nadesan & Shah, 2020). Incorporating the oral communication skills into the Malaysian CEFR aligned curriculum does not necessarily result in enhanced ease of learning for the students (Kaur, 2022). Given that speaking is a component of the Malaysian Certificate of Education (SPM) examination, the emphasis on grammar-based instruction and writing proficiency necessitates a shift towards a comprehensive approach encompassing all four language skills - speaking, listening, reading, and writing. The initiative undertaken by the MOE to revise the curriculum can be viewed as a double-edged sword. In addition to the restructuring of the educational system to promote a more comprehensive approach, it also affords teachers and students the chance to explore novel approaches for language learning, specifically in speaking, which is the main focus of this study.

One possible way to realise this is to reconceptualise assessment by combining assessment for learning (AfL) and assessment of learning (AoL) (Hosseini & Nimehchisalem, 2020). Utilising self-assessment, as a synergy of both AfL and AoL, is an avenue to achieving this goal. This study's undertaking is driven by past studies that have shown that secondary school students can effectively engage in self-assessment using rubrics (Andrade, 2019; Guzman, 2022) and the efficacy of self-assessment in its capacity to benefit students (Su, 2020; Vasileiadou & Karadimitriou, 2021; Xiao & Yang, 2019) in improving their performance in second language learning (Anh et al., 2022; Avilez & Larenas, 2020) and fostering agency, autonomy, and self-regulated learning opportunities (Hosseini & Nimehchisalem, 2021; Marzuki et al., 2020; Masruria & Anam, 2021; Mohamed Jamrus & Razali, 2019; Sintayani & Adnyayanti, 2022; Xiao & Yang, 2019).

While most Malaysian-based studies on self-assessment and speaking have focused on higher institution students (Ahmad Bukhari, 2018; Rahman et al. 2025; Razali & Abdul Latif, 2019), there is little focus on secondary school ESL students, a demographic that faces challenges in speaking (Kaur, 2022; Sintayani & Adnyayanti, 2022), which has been significantly overlooked. This study, therefore, aims to bridge the gap by examining its application and effectiveness in the Malaysian ESL secondary school context. In order to maximise the effective utilisation of self-assessment and facilitate students to derive benefits
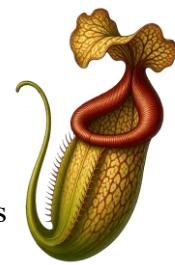
from self-assessment, rubric training sessions (Su, 2020) encompassing the use of exemplars (Hawe et al., 2021) assume a pivotal role in the present study.

The purpose of this study is to i) investigate the effect of employing the Speaking Self-assessment Instrument (SSAI) (i.e., a rubric specifically designed for this study) on the speaking performance of students, and ii) examine the effect of the SSAI on students' grammar, vocabulary and communicative competency. As such the research questions of this study are as follows:

1. Does the SSAI as a learning tool improve the speaking performance of students?
2. To what extent does the implementation of the SSAI improve the students' grammar, vocabulary and communicative competence?

To answer the first research question, it is hypothesised that:

H_{01}**:** There is no relationship between the use of the SSAI as a learning tool and students' performance in speaking.
H_{02}: There is a positive relationship between the use of the SSAI as a learning tool and students' performance in speaking.

LITERATURE REVIEW

This section presents the framework that guides this study and addresses literature related to self-assessment followed by the introduction of rubrics as a form of self-assessment. This is followed by an overview of rubric training and the use of exemplars with close reference to relevant literature which resonates the present study. This section concludes with a review of literature pertinent to students' speaking performance and linguistic output.

Self-Assessment to Promote Self-Regulated Learning (SRL)
Self-assessment is a process whereby students assess the quality of their cognitive processes and behaviours throughout learning, while identifying possible ways to enhance their knowledge and skills (McMillan & Hearn, 2008; Panadero et al., 2016; Panadero & Jonsson, 2020; Rolheiser & Ross, 2001). By engaging in self-assessment, students evaluate their own work to enhance performance by recognising gaps between actual and anticipated outcomes (McMillan & Hearn, 2008; Yan et al., 2022) in any activities that involve acquiring knowledge and skills (Goetz et al., 2013; Panadero & Jonsson, 2020). In this study, self-assessment is highlighted as it promotes SRL (Benraghda et al., 2022) wherein student agency in taking responsibility and autonomy of their own learning is emphasised (Piccardo & North, 2019). This emphasis is backed up by past research which show that students who practice SRL are accountable in consistently monitoring their learning progress and taking necessary actions to achieve their learning goals which in turn leads to enhancement in their academic performance (Andrade, 2019; Panadero & Jonsson, 2020).

In the context of this study, self-assessment was introduced to the student-participants with close reference to Vygotsky's (1978) sociocultural theory. As the student-participants were not familiar with self-assessment, they were provided with training and exemplars on how to self-assess themselves through the process of scaffolding by the researcher-participant via familiarisation to the SSAI, a rubric for self-assessment. This scaffolding process aids the students to transition from being teacher-dependent to being independent learners (Carless,

2007) in self-assessing their speaking performance via both the rubric training and use of exemplars. Allal (2016) defines this as co-regulation in which students acquire self-regulation strategies by engaging with a more proficient individual, such as a teacher or an advanced peer.

Rubrics as a Tool for Self-Assessment

In this study, the SSAI (i.e., a rubric) is developed as a tool for self-assessment in speaking. This is motivated by the capacity of rubrics in assisting students to comprehend the disparity between their performance and expected outcomes or targeted goals (Panadero & Jonsson, 2020) based on its features which explicitly outline specific tasks and the standards for various levels of performance (Brookhart, 2013). Rubrics are also suitable for the age group (i.e., 16 years old) of the participants in this present study (Andrade & Brown, 2016; Vasileiadou & Karadimitriou, 2021) and have been proven to generate beneficial impacts on learning and academic achievement (Benraghda et al. 2022), specifically in relation to speaking (Panadero et al., 2023; Su, 2020; Yufrizal et al., 2020).

Past studies have proven that self-assessment through the use of rubrics improves students' speaking performance because the act of self-assessing requires the students to evaluate their learning process, product, and development (Sintayani & Adnyayanti, 2022). The explicit criteria in a rubric gives the students a clear idea of what is expected of them in a speaking test which results in building their self-awareness and identification of aspects that they need to improve to enhance their speaking performance (Smyth & Carless, 2020). Additionally, Li and Zhang (2020) posit that self-assessment, which they refer to as internal assessment, could be used as a supplement to external assessment methods like language tests and teacher evaluations. Their meta-analysis between the correlation of self-assessment and language performance showed that specific rubric criteria and training improved the correlation strength.

While a number of research show positive correlations between self-assessment and speaking performance, studies providing empirical evidence or measurable data on students' actual improvement in speaking skills are notably scarce. One significant study that gauges students' speaking performance in terms of accuracy and fluency is by Huang and Gui (2015) who found that rubrics improved EFL learners' discourse length, organisation, and flexibility but not accuracy in their speaking performance. Hung's (2019) study, similarly, shows that EFL students' speaking performance improved in regards to fluency but was less pronounced in grammar and vocabulary. However, instead of using rubrics, the students in Hung's study were introduced to self-assessment in the form of a checklist (i.e., a form containing guiding questions). Another study focussing on speaking fluency by Santos and Ramirez-Avila (2022), echoes the same result wherein the students in their study showed better performance in terms of the number of words they spoke, less hesitations, and decreased number of repetitions and corrections. In the present study, the students' speaking performance was examined by analysing both their fluency and accuracy in terms of grammar, vocabulary and communicative competence with close reference to the SPM speaking scales provided by the MOE (MOE, 2021).

Rubric Training and Exemplars

In relation to the use of rubrics, it is imperative for educators to provide students with proper instruction and guidance in their utilisation (To et al., 2021). Rubric training is vital in order to optimise the efficacy of rubric utilisation (Andrade, 2019; Brookhart, 2013; Vasileiadou & Karadimitriou, 2021; Su, 2020). This study utilised the training technique adapted from Su's

(2020) SIB scheme, which was derived from Patri's (2002) work. The acronym SIB is comprised of three phases – a) studying rubrics (S), b) identifying performance features (I) and c) bridging the gap (B) (Su, 2020). In phase S, the student-participants in this study engaged in the examination of the SSAI focussing on the band descriptors for each criteria prior to self-assessing themselves. This involved scaffolding (Vygotsky, 1978) and feedback (Hattie & Temperly, 2007) from the teacher-researcher's part to help students comprehend and interpret the criteria and descriptors in the SSAI for accurate application in their self-assessment.

Exemplars were utilised to help student-participants translate key features from the SSAI into actual performance in phase I. Exemplars are commonly defined as products or processes that represent specific levels of quality or competence (Sadler, 1987; To et al., 2021), often the ideal and least favourable practices, assisting students in understanding the criteria and standards required for assessment tasks (Chong, 2020; Chong & Ye, 2020; Newlyn, 2013). In the current study, sample videos of students performing speaking tasks (i.e., in three different levels) taken from YouTube, along with their band scores (i.e., A1-A2, B1-B2, and C1), were utilised as exemplars to assist students to better grasp what is expected of them at each level of performance, leading to more accurate self-assessment. Following this, the student-participants discussed among themselves on how to close the gap between their current level and the exemplar level which was done in phase B.

Smyth and Carless's (2020) study suggests that combining exemplars with rubrics and self- or peer-assessment strategies significantly enhanced students' understanding of academic standards which led to self-monitoring. The authors postulate that a combination of exemplars and rubrics improve academic performance and promote academic self-regulation. This perspective is echoed by Hawe et al. (2021), who posit that exemplars are most effective when educators view self-regulation as their primary goal. They posit that exemplars can support students in transferring assessment and productive skills to future tasks.

Speaking Performance and Linguistic Output
In this study, the relationship between the students' linguistic output in spoken form in terms of their grammar, vocabulary and communicative competence from their pre-test and post-test speaking scores are investigated to elucidate the variations suggesting improvement on the students' speaking performance.

Ur (2001) describes grammar as language structure rules which involve arranging words, phrases, and sentences to express meaning which affect how language is used in communication. According to Kang and Yan (2018), in the context of assessing speaking proficiency, grammar is measured via two key aspects which are accuracy (e.g., error-free in the use of articles, prepositions, singular/plural, and subject-verb agreement (SVA) and complexity (e.g., the production of more complex sentences with dependent clauses). In this study, aligned with the two key aspects mentioned and with close adherence to the SPM assessment speaking scales (MOE, 2021), there were three main elements that influenced grammar assessment:

1. The students' ability to communicate accurately
2. The students' use of accurate and appropriate grammatical structure
3. The students' competence of linking utterances using appropriate cohesive devices such as conjunctions, relative pronouns and ellipsis

In regards to vocabulary assessment, there were two components assessed which were vocabulary size and vocabulary depth. This was adopted from the study by Enayati and Derakhshan (2021) which investigated the influence of vocabulary size and depth on predicting the speaking proficiency of second language learners. Their study revealed a significant correlation between higher vocabulary size and better speaking abilities, enabling learners to express themselves more clearly and effectively. Moreover, vocabulary depth, signifying an in-depth understanding of word meanings, enabled students to articulate complicated ideas more successfully.

In assessing communicative competence, with close reference to the SPM speaking assessment scales, there were three factors assessed. First, the students' conversational sustainability. Second, the students' communication clarity and third, the students' ability to convey their speaking partner's message (MOE, 2021). This is in line with Harding et al.'s (2023) concept of test-taking competence as a type of communicative competence. According to Harding et al. (2023), in a test-taking context, communicative competence encompasses learners navigating their cognitive skills (i.e., phonological, lexical, morphological and synthetic knowledge) as well as social interactional abilities (i.e., task achievement, appropriateness, turn-taking, and promptness of response) in speech production.

MATERIALS AND METHODS

This study employed a quasi-experimental one-group pre-test-post-test design. The independent variable in the study is the utilisation of the SSAI, the dependent variable is the students' speaking performance, and the moderating variable is the varying levels of student proficiency.

Purposive sampling was used for participant selection in this study. Three public secondary schools in Sandakan, Sabah, Malaysia were selected for their accessibility and practicality as the researcher was located within the same vicinity. A permission application was submitted to the Educational Research Application System (eRAS 2.0) and the Sabah State Education Department prior to the conduct of field work. Upon approval, assent was obtained from the respective schools' principals and students. 151 Form 4 students from six classes volunteered to participate in this study. The students' English language proficiency level ranged from lower-intermediate (58.94%), intermediate (39.74%) and upper-intermediate (1.32%). The data were obtained from the students' pre-test speaking test scores. All of the students were 16 years old. Their first and most spoken language is the Malay language. The students' names, the school's name and any other identifying information were removed to preserve anonymity. In the findings, the students are represented by alphanumerical codes.

Instrument
The SSAI is a student's self-assessment analytical rubric consisting of ten criteria involved in evaluating students' speaking performance adapted from studies by Babaii et al. (2015), and Mori and Pell (2019), and the SPM English – Speaking assessment scales (Ministry of Education, 2021). The ten criteria are i) grammar, ii) vocabulary, iii) communicate competence, iv) fluency, v) pronunciation, vi) topic management, vii) confidence, viii) organisation, ix) strategy use, and x) time management. Each criterion is accompanied by its own set of can-do descriptors (Council of Europe, 2020; Piccardo & North, 2019) and
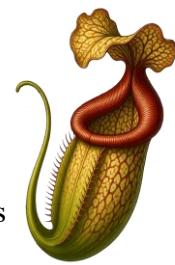
corresponding band levels. The SSAI is presented in both the English and Malay languages (See Appendix C).

Data Collection

The students were given a speaking pre-test prior to the intervention (i.e. the use of the SSAI). The speaking pre-test was conducted at school and students participated in pairs. They were assessed and video recorded. Video recording was necessary for the second assessor (i.e., an experienced teaching English as a Second Language (TESL) teacher of 11 years residing in Kota Kinabalu, Sabah) to be able to assess the students at his convenience. The presence of a second assessor was crucial to avoid researcher bias and ensure the reliability of the test scores. Pearson's Correlation was utilised to check the degree of linear correlation between the researcher and second assessor's scores (Goos & Meintrup, 2016). Given the time constraints and assessment capacity of the second assessor, 15 students representing 10% of the total participants, were randomly selected and evaluated for their speaking pre- and post-tests by the second assessor. Analysis via the Statistical Package for the Social Sciences (SPSS) version 29 demonstrated a strong positive correlation ($r = .974$, $p < .001$), signifying a high degree of agreement between the assessors (Goos & Meintrup, 2016) for the speaking pre-test scores. For the speaking post-test scores, the analysis revealed a substantial positive connection ($r = 0.987$, $p < .001$) indicating a strong consensus between the assessors' scores. The speaking tests, in adherence to the SPM speaking test format, consisted of three parts – interview about self, individual long-turn task and discussion task. The students were assessed and video recorded in pairs by the researcher-participant at their respective schools.

Following the speaking pre-test, the students participated in four SSAI familiarisations and rubric training sessions (i.e., a total of two English language lessons, each lasting for a duration of 60 minutes, scheduled once each week over a span of four weeks). The rubric training session encompasses:

1. The introduction to the SSAI (i.e., explanation of the descriptors and bandscores to the students
2. Presenting them with exemplar videos and collectively assessing and rating the exemplar candidates from the videos
3. Students self-assessing themselves based on their respective speaking test videos
4. Feedback and discussion regarding the SSAI, students' self-assessment accuracy and their speaking performance

The speaking post-test was administered after the intervention. The questions presented during the speaking post-test differed from those in the pre-test speaking test, however they adhered closely to the framework specified for the SPM speaking test and were adapted from the questions in the speaking pre-test.

Data Analysis

The students' test scores from the pre and post-test speaking tests were analysed using a mixed ANOVA (Analysis of Variance) model via the SPSS version 29. To answer the first research question, the mixed ANOVA was utilised to determine whether there is a significant interaction between the proficiency level and the effect of using the SSAI as well as whether the SSAI improves performance overall. The between-subjects factor was the students' proficiency levels (i.e., lower-intermediate, intermediate, upper-intermediate). The within-subjects factor was the use of the SSAI and the comparison between the pre-test and post-test speaking test scores.

To analyse data of the second research question, the SPM speaking assessment scales were adapted and utilised to examine the students' language production. The SPM speaking assessment scales encapsulate three language components - grammar, vocabulary, and communicative competence, and are structured as an analytic rubric (MOE, 2021). There were in total eight pre- (i.e., four videos) and post-test videos (i.e., four videos) analysed based on the selection of student-participants that showed significant difference in the pre-test and post-test scores. Only four sets of students' pre- and post-test videos were selected due to the rigorous and time-consuming process in transcribing the speaking test videos. The students' linguistic output were assessed based on these criteria – grammar, vocabulary, and communicative competence (i.e., as listed in the SPM speaking assessment scales).

RESULTS

This section outlines the findings of the study concerning students' pre-test and post-test speaking scores and the students' linguistic output.

The Pre and Post-Test Scores

Table 1
Students' Pre-test and Post-test Speaking Test Scores

	Proficiency level	Mean	Std. Deviation	N
Pre-test Score	Lower-intermediate	4.29	2.464	89
	Intermediate	13.50	2.931	60
	Upper-intermediate	22.50	.707	2
	Total	8.19	5.474	151
Post-test Score	Lower-intermediate	7.11	4.152	89
	Intermediate	16.53	4.612	60
	Upper-intermediate	23.00	1.414	2
	Total	11.07	6.452	151

Based on Table 1, the pre-test scores revealed that students at the upper-intermediate level achieved the highest average score (Mean = 22.50), whereas those at the lower-intermediate level attained the lowest average score (Mean = 4.29). The standard deviations revealed diversity in scores across all levels, with the intermediate group exhibiting the highest standard deviation (2.931), indicating greater score variability among these students. The post-test results demonstrated considerable improvement across all proficiency levels, especially for the lower-intermediate group, which increased from a mean of 4.29 (pre-test) to 7.11 (post-test). The intermediate group exhibited an increase in the mean score from 13.50 (pre-test) to 16.53 (post-test). The upper-intermediate group experienced a marginal rise in mean score from 22.50 (pre-test) to 23.00 (post-test). The overall mean score for pre-tests was 8.19, which rose to 11.07 in post-tests, indicating an improvement across all students, with the greatest impact in the intermediate and upper-intermediate groups.

Table 2
Tests of Within-subjects Effects

Source		F	Sig.	Partial Eta Squared
Pre-test and post-test	Sphericity Assumed	5.723	.018	.037
Pre-test and post-test * Proficiency level	Sphericity Assumed	.488	.615	.007

Table 2 shows that the SSAI had a significant influence on speaking performance (F-value = 5.723, p-value = 0.018) pre- and post-test. This implies that the SSAI considerably enhanced students' speaking performance between pre- and post-test. Partial Eta Squared of 0.037 indicated a small influence (Cohen, 1988). Pre-test versus post-test and student proficiency did not interact. The F-value for this interaction was 0.488 and the p-value was 0.615, which exceeded the 0.05 statistical significance threshold, indicating that the null hypothesis was not rejected. The difference between pre-test and post-test scores (i.e., the effect of the SSAI on speaking performance) did not change significantly across proficiency levels, showing that the SSAI had a similar effect. The tiny effect size (Partial Eta Squared = 0.007) supports this.

Table 3
Tests of Between-subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	7158.693	1	7158.693	374.783	<.001
Proficiency Level	6916.528	2	3458.264	181.052	<.001

Table 3 shows the F-statistic for the intercept is 7158.693, accompanied by a p-value of less than 0.001. This signifies a highly significant intercept (Field, 2013), demonstrating that there was a significant overall mean for the dependent variable, specifically speaking performance as measured by speaking test scores. The F-statistic is 374.783, accompanied with a p-value below 0.001. This indicates substantial disparities (Field, 2013) in average scores among the various proficiency levels. The findings indicate that the students' proficiency level significantly influenced their average scores.

Table 4
Post Hoc Comparisons across Students' Speaking Performance and Levels of Proficiency

Multiple Comparisons

Tukey HSD

(l) Proficiency Level	(j) Proficiency Level	Mean Difference (l-j)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Lower-intermediate	Intermediate	-9.31*	.516	<.001	-10.54	-8.09
	Upper-intermediate	-17.05*	2.210	<.001	-22.28	-11.82
Intermediate	Lower-intermediate	9.31*	.516	<.001	8.09	10.54
	Upper-intermediate	-7.73*	2.221	.002	-12.99	-2.47
Upper-intermediate	Lower-intermediate	17.05*	2.210	<.001	11.82	22.28
	Intermediate	7.73*	2.221	.002	2.47	12.99

Table 4 demonstrates notable distinctions across the three proficiency levels. The intermediate group demonstrated superior performance compared to the lower-intermediate group (mean difference = 9.31, p < 0.001; 95% CI: 8.09 to 10.54). Additionally, the upper-intermediate group achieved higher scores than the lower-intermediate group (mean difference = 17.05, p < 0.001; 95% CI: 11.82 to 22.28). The upper-intermediate group outperformed the intermediate group, with a mean difference of 7.73 (p = 0.002; 95% CI: 2.47 to 12.99).

Linguistic Output
This section will be discussed in three parts – a) grammar, b) vocabulary and c) communicative competence.

Grammar
In terms of grammar, it was found that lower-intermediate and intermediate students, despite making grammatical errors in their sentence structures, managed to exhibit the usage of conjunctions such as 'and' and 'because' and link their utterances better in their speaking post-test test with less prompting from the assessor. For instance, as shown in Table 7, for Speaking Part 2 in which S104 had to talk about a celebration that he had celebrated recently, the assessor had to prompt S104 a few times to elaborate on his answer:

Table 7
Transcript of Speaking Pre-test (S104)

Test (Part)	Speaker	Transcription
Pre-test (Speaking Part 2)	S104 (Lower-intermediate)	Last? Uh [long pause] can same [referring to his speaking partner's response]? Happy birthday.
	A	Okay. Did you have a good time?
Appendix A	S104	Yes, I enjoy myself.
	A	Okay, why?
	S104	Because many- peo- my friend [long pause]
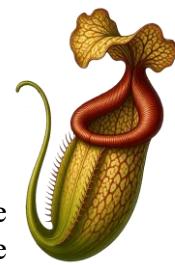
As shown in Table 8, for Speaking Part 2 in the speaking post-test, on addressing the point of 'Why school holidays are important' (see Appendix B), S104 managed to answer the question without prompting from the assessor and utilised 'because' and 'and' to form his utterance:

Table 8
Transcript of Speaking Post-test (S104)

Test (Part)	Speaker	Transcription
Post-test (Speaking Part 2) Appendix B	S104 (Lower-intermediate)	School holiday are important because uh student have uh spend time with family and friend.

Even though S104 managed to link his ideas, it is evident that the utterance is grammatical inaccurate in terms of his usage of the noun 'holiday' which should be used in the plural form to indicate a recurring or general event and the misuse of the auxiliary verb 'have' which should function as a lexical verb that requires a noun complement (i.e., time) followed by the 'to-infinitive' (Huddleston & Pullum, 2002). It can be postulated that S104 has not mastered the grammar rules of the language but has attempted the usage of conjunctions to better his speaking performance. This could possibly be influence by the SSAI which specifies the use of connectors to link sentences under the criterion of 'Organisation' (see Appendix C). This corroborates Hung's (2019) and Huang and Gui's (2015) findings which have shown improvement in terms of the participants' discourse length and inaccuracy in their grammar.

For an intermediate student like S31, prompting was less required. As shown in Table 9, when asked about the person that she admired, she was able to comprehend the task and responded to it. From her utterances, it could be conceived that these were chunks of her ideas which were not coherently connected. Even so, it can be seen that she attempted to connect her ideas by using the connector 'so' twice.

Table 9
Transcript of Speaking Pre-test (S31)

Test (Part)	Speaker	Transcription
Pre-test (Speaking Part 2) Appendix A	S31 (Intermediate)	I have a person that I admire very much. She is a peers of mine. I used to be in XYZ before I transferred here so she's a very good person for me. I was struggling so much during my time in boarding school so she helps me a lot.

After the intervention, it could be seen in Table 10 that S31 was more verbatim and utilised more connectors and cohesive devices to connect her utterances. S31 was also capable of using the relative clause 'that' to modify the noun 'teacher' and the complementiser 'that' which introduced a complement clause (i.e., that I can be confident) but has no function within it (Huddleston & Pullum, 2002). However, similar to the previous student, S104, grammatical errors were still evident in her speaking post-test. For instance, S31 committed an auxiliary verb error, which is an SVA error (Murphy, 2019) in which she used 'have' instead of 'has' to refer to her singular teacher. In line 4, nevertheless, she managed to use the verb 'has' correctly. There are also other instances of SVA errors as indicated by the underlined texts in Table 10.

Table 10
Transcript of Speaking Post-test (S31)

Test (Part)	Speaker	Transcription
Post-test (Speaking Part 2) Appendix B	S31 (Intermediate)	I have a teacher that I admire. I met this teacher during my Add Maths class on Form 4. She is teacher Y. She's uh- she have been teaching me for a year and one month now and she's- I personally think she's an admirable teacher because she has a bold personality it always convince me that I can be confident even though I- I don't really master Additional Math subject at all and she always give us motivation…encourage us to make sure we study properly, to uh think of our future and I think it's a good thing for us to have a teacher that give us reminder.

Vocabulary

Mos of the students could also demonstrate the use of sufficient vocabulary to talk about familiar topics better in the post-test. This is evident in students from the lower-intermediate group. For instance, as shown in Table 11, it can be seen that when S152 was questioned about the importance of having a good role model, he struggled to talk about a person that he admired and was not able to answer the question:

Table 11
Transcript of Speaking Pre-test (S152)

Test (Part)	Speaker	Transcription
Pre-test (Speaking Part 2) Appendix A	S152 (Lower-intermediate)	I am going to talk about a person I admire. Who this person…this person is…my…[long pause]
	A	It's okay, you can move to the next question
	S152	…if I think it's important for people to have role model…yes…this is because…[long pause]
	A	Okay, why do you think it's important to have a good role model?
	S152	[Long pause; places his left hand on his head and shakes his head]

In the speaking post-test, based on Table 12, S152 was able to talk about what he did during the school holidays without prompts. S152 addressed all the points required in Part 2 of the speaking test. However, it was evident that S152 committed grammatical errors in terms of his tenses (e.g., I spend my..; I go out..;) whereby he used the present tense to indicate past actions (Murphy, 2019).

Table 12
Transcript of Speaking Post-test (S152)

Test (Part)	Speaker	Transcript
Post-test (Speaking Part 2) Appendix B	S152 (Lower-intermediate)	I am going to talk about what I did during the school holidays. Where did I spend my holidays, I spend my holidays at home but sometime I go out with my brother, my friend or brother to spend more time and go eat some- somewhere. What I did there. At home I always playing game Mobile Legend such as Mobile Legend or maybe sometime go to Bataras to play some game and buy something or food that we

	want. If I enjoy myself, yes cause- because I can spend more time with my brother and friend and I'm not only stay at home a lot.

S47, who was also a lower-intermediate student, produced more utterances in her speaking post-test (see Table 14) if compared to her speaking pre-test (see Table 13). When asked to describe a person that she admired, S47 only addressed three out of the four points required (i.e., 'Who the person is', 'What this person looks like', 'Why you admire this person', and 'If you think it is important for people to have good role models - why / why not?') (see Appendix A) suggesting limited vocabulary to express her ideas. A few grammatical errors were also present as underlined. In regards to word choice, S47 incorrectly used the word 'beauty' (noun) instead of 'beautiful' (adjective) to describe the person that she admired (Murphy, 2019).

Table 13
Transcript of Speaking Pre-test (S47)

Test (Part)	Speaker	Transcript
Pre-test (Speaking Part 2)	S47 (Lower-intermediate)	My…my person is uh SX because she person is looks like beauty and be nice. She um…very clever and uh that's all *lah*.
Appendix A		

For the speaking post-test (see Table 14), S47 managed to address all the points as mentioned prior including point number four which is 'If you think it is important for people to have good role models (why/why not?)' (see Appendix B). S47 understood the task and were able to convey her ideas even with limited vocabulary as there were no supporting details and elaboration to her answers given.

Table 14
Transcript of Speaking Post-test (S47)

Test (Part)	Speaker	Transcript
Post-test (Speaking Part 2)	S47 (Lower-intermediate) A	Uh who the teacher is T1 because T1 looks like uh pretty girl and quite nice. Um I admire the teacher because uh she uh she…she clever Mm-hmm…
Appendix B	S47	Uh…I think it's important for teacher to be good role models uh *ya* (yes) because she can improv- improve me to speak uh English. Yes, that's all *lah*.

Communicative Competence
In terms of communicate competence, changes were evident in Part 3 of the speaking test whereby students were more engaging towards each other in discussing the points given without the need for prompts from the assessor. This was less evident with students in the lower-intermediate group and more conspicuous with students in the intermediate and upper-intermediate groups. For instance, based on Table 16, in the speaking pre-test Part 3 whereby the students had to discuss about the usage of mobile phones with each other (see Appendix A), the assessor had to intervene and mediate the discussion between S31 and S32 to keep the conversation flowing.

Table 16
Transcript of Speaking Pre-test (S31 and S32)

Test (Part)	Speaker	Transcription
Pre-test (Speaking Part 3) Appendix A	S31 (Intermediate)	On the other hand, using mobile phones can be- can make students most of the time at disadvantage because they tend to spend more day time on phones rather than in real life or to study or maybe to even help their parents at home because they think- because they make phones as their essential of life like the main priority of their life. [nods in silence]
	A	How about you?
	S32 (Intermediate)	Well, for me the advantage of using mobile phones is definitely to study…however the disadvantage is people tends to use it for cyber-bullying, harassing people online and that's not a very good thing and people use mobile phone. *Ya* (yes)…[nods in silence]
	A	Anything else to add?

In contrast, in the speaking post-test, it was apparent that there were turn-taking occurrences between S31 and S32 in their discussion of why people use social media (see Appendix B part 3). From Table 17, it could be seen that after elaborating on her points, S31 asked S32 if she had anything to add to the discussion. S32, in reply, agreed with the points given by S31 and added further justification to support her point, which showed her synthetic knowledge (Harding et al., 2023) of the topic discussed. It could be postulated that both S31 and S32 showed the capacity to sustain the conversation which is one of the factors assessed under the criterion of communicative competence (MOE, 2021).

Table 17
Transcript of Speaking Post-test (S31 and S32)

Test (Part)	Speaker	Transcript
Post-test (Speaking Part 3) Appendix B	S31	For me, I also use social media to connect with people because I've experienced um pen pals and online friends through my social media like recently uh I just found a friend from Germany and also from United Kingdom. They have been very good to me. It's like we feel connected as we have similar interests….it proves that people around the world can connect from just one simple application. Do you have anything else?
	S32	Yes um I also agree with you have said before because you know recently after Covid 19 people has been really disconnected to each other so um…

DISCUSSION

This study aimed to examine the effectiveness of the SSAI as a learning tool on Malaysian upper secondary school students' speaking performance, specifically on their linguistic output which encompassed grammar, vocabulary and communicative competence. The results indicate a significant positive effect on the usage of the SSAI on overall speaking performance across students from different levels of proficiency, rejecting the null hypothesis, particularly significant for lower-intermediate students and intermediate students.
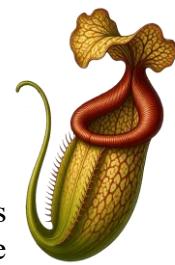
In relation to grammar, basic connectors such as 'and' which is also referred to as additive conjunction (Halliday & Hasan, 1976) and 'because', a causal conjunction, were found in the students' utterances more in their post-test speaking test. This suggests that the students were able to show their ability to link, relate and reinforce their ideas, produce longer discourse and require less prompting by the assessor. An intermediate student, for instance, was also able to use the relative clause, complementiser, and cohesive devices accurately in her utterances in the post-test. However, there were also grammatical inaccuracies evident in the students' utterances – SVA error (e.g., the usage of 'have' to indicate a singular entity), auxiliary verb misuse (e.g., 'have' instead of 'to-infinitive') and incorrect noun form (e.g., 'holiday' instead of 'holidays'). These findings corroborate with Hung (2019) and Huang and Gui's (2015) studies wherein self-assessment resulted in the enhancement of the EFL learners' discourse length and organisation but did not improve grammatical accuracy.

In regards to vocabulary, the students demonstrated the ability to use sufficient vocabulary to discuss familiar topics more effectively in the post-test. This is significantly evident among students in the lower-intermediate group. For example, a lower-intermediate student who struggled with a topic in the pre-test (i.e., to talk about a role model) and could not answer questions without prompts, was able to speak about school holidays in the post-test, addressing all points, despite some grammatical errors in tenses (e.g., using present tense for past actions). This is also shown by another lower-intermediate student who managed to address all the discussion points albeit with limited vocabulary and word choice error (e.g., 'beauty' instead of 'beautiful'). This suggests that the students were able to understand the topic that they were given and showed ability to reason their claims, supporting the claim that vocabulary size and depth are determinants of the students' speaking ability (Enayat & Derakhshan, 2019). Both students also showed better fluency whereby there was an evident increase of word utterances and less hesitations, although with grammatical errors, which aligns with Santos and Ramirez-Avila's (2022) study suggesting that self-assessment helps in enhancing the students' speaking fluency.

Analysis of post-test speaking data also revealed the students' improvement in communicative competence, particularly among intermediate and upper-intermediate students. In the pre-test, for Part 3 of the speaking test, majority of the students needed intervention by the assessor to mediate their discussion. For instance, when talking about the advantages and disadvantages of mobile phones, students tend to merely state their points individually without asking opinions from their respective partners. Additionally, the students did not comment on their partner's suggestions resulting in more prompting by the assessor. In contrast, the students exhibited improved turn-taking and a more sustained conversation in their post-test. S32, for instance, exhibited synthetic knowledge (Harding et al., 2023) wherein she not only agreed to S31's statement, but she was also able to add information to the conversation. The use of the SSAI is believed to have supported the students in task performance due to its explicit descriptor such as 'I can work well and negotiate towards an outcome with my partner' (i.e., Bandscore/level 5) which indicates the need for negotiation from both candidates and with the support of the exemplar videos demonstrating how to achieve such level with reference to real students' speaking performance at this level. This is in line with Smyth and Carless (2020) wherein they suggested that combining exemplars with rubrics enhance students' understanding of academic standards which in turn leads to improvement in academic performance (Panadero et al., 2023).

The primary limitation of the study is the lack of control groups. The students included in the study were drawn from the classes of participating teachers as designated by the school administration. It was not possible for the students to be classified into control and experimental groups due to an uneven distribution of proficiency levels among them. In addition, the student enrolment in each class varied. In the absence of control groups in this study, repeated measures Analysis of Variance (ANOVA) was utilised (Creswell & Creswell, 2018) to eliminate variability attributed to individual differences. The presence of control groups would provide further evidence of the effectiveness of the SSAI in enhancing students' speaking performance. Moreover, a delayed post-test to assess any significant improvement in students' speaking performance was not conducted due to time constraints. Findings from Hung (2019) and Su's (2020) study indicate that both the learners' assessment skills and speaking performance improved with time. An extended study period could be taken into consideration for future research in this area of study.

CONCLUSION

This study presented empirical data that demonstrate the effectiveness of self-assessment in the improving speaking performance of Malaysian ESL upper secondary school students. The post-test speaking test results showed enhanced linguistic features such as the usage of connectors, cohesive devices and more topic-related vocabulary with the most pronounced effect observed in lower-intermediate and intermediate students. Enhancement in communicative competence is prominently exhibited by the upper-intermediate students who showed improvement, although with modest increase, in turn-takings and engagement in discussions with less prompts by the assessor. The overall findings of this study suggest that the employment of self-assessment is directly linked to the tangible improvements in students' linguistic output. This study implies that self-assessment can be employed in the secondary school ESL classrooms to help students take agency of their learning, specifically in speaking by helping them to self-regulate their learning via the use of SSAI (i.e., descriptors in the rubric) to bridge the gap between their current speaking performance and their targeted goals.

REFERENCES

Ahmad Bukhari, N. I. (2018, October 21-November 1). *CEFR gauging of ESL learners' speaking skills in Malaysia: A conceptual paper*. International Conference on Ummah ICU. Kelantan, Malaysia.

Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In D. Laveault & L, Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259-274). Springer.

Andrade, H. L. (2019). A rritical review of research on student self-assessment. *Front. Educ. 4*(87), 1–13.

Andrade, H. L. & Brown, G. T. L (2016). Student self-assessment in the classroom. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 319-334). Routledge.

Anh, K. H., Dong, M. H. & Trang, N. H. (2022). An investigation into English majored students' self-assessment of their speaking performance. *International Journal of Instruction, 15*(3), 191-208. https://doi.org/10.29333/iji.2022.15311a
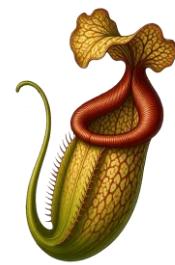
Avilez, A. P. & Larenas, C. D. (2020). Enhancing English speaking skills through self-assessment-based action research. *International Journal of Action Research, 16*(2), 132-152. https://doi.org/10.3224/ljar.v16l2.04

Babaii, E., Taghaddomi, S. & Pashmforoosh, R. (2015). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*. 1-27. https://doi.org/10.1177/0265532215590847

Benraghda, A. (2022). Self-assessment as a self-regulated learning approach in English oral presentations: College students' choices and perceptions. *Cogent Education, 9*(1), Article e2123472. https://doi.org/10.1080/2331186X.2022.2123472

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). Longman.

Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57-66. https://doi.org/10.1080/14703290601081332

Chong, S. W. (2020). The use of exemplars in English writing classrooms: From theory to practice. *Assessment & Evaluation in Higher Education, 44*(5), 748-763. https://doi.org/10.1080/02602938.2020.1844196

Chong, S. W., & Ye, X. (2020). *Developing writing skills for IELTS: An exemplar-based instructional approach*. The Education University of Hong Kong. https://www.eduhk.hk/cle/resources/cep/tdg-master/examplar.html

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://www.coe.int/lang-cefr

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. (5th ed.). Sage Publication.

Enayat, M. J. & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, 102521. https://doi.org/10.1016/j.system.2021.102521

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). Sage Publication.

Franchis, C. & Mohamad, M. (2023). Implementation of ESL assessment in Malaysian schools: A systematic literature review paper. *International Journal of Academic Research in Progressive Education & Development, 12*(2), 2311-2326. http://dx.doi.org/10.6007/IJARPED/v12-i2/17537

Goetz, T., Nett, U. E. & Hall, N. C. (2013). Self-regulated learning. In N. C. Hall & T. Goetz (Eds.), *Emotion, motivation, and self-regulation: A handbook for teachers* (pp. 123-158). Emerald.

Goos, P. & Meintrup, D. (2016). *Statistics with JMP: Hypothesis tests, ANOVA and regression*. Wiley.

Grainger, P. R., Heck, D. & Carey, M. D. (2018). Are assessment exemplars perceived to support self-regulated learning in teacher education? *Frontiers in Education, 3*(60), 1-9. https://doi.org/10.3389/feduc.2018.00060

Guzman, R. E. B. (2022). *Applying self-assessment to improve speaking skills in a secondary school* [Master's thesis, Universidad Casa Grande]. Repositorio UCG. http://dspace.casagrande.edu.ec:8080/bitstream/ucasagrande/3460/1/Tesis3578BORa.pdf

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. Longman.
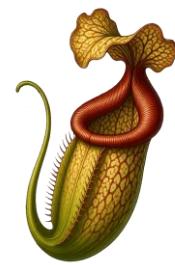
Harding, L., Macqueen, S. & Fill, J. (2023). Assessing communicative competence. In M. Kanwit & M. Solon (Eds.), *Communicative competence in a second language: Theory, method, and applications* (pp.187-207). Routledge.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hawe, E., Dixon, H., & Hamilton, R. (2021). Why and how educators use exemplars. *Journal of University Teaching & Learning Practice, 18*(3). https://doi.org/10.53761/1.18.3.10

Hilden, R., Oscarson, A. D., Yildirim, A., & Frojdendahl, B. (2022). Swedish and Finnish pre-service teachers' perceptions of summative assessment practices.*Languages, 7*, 10, 1-23. https://doi.org/10.3390/languages7010010

Hosseini, M. & Nimehchisalem, V. (2021). Self-Assessment in English language teaching and learning in the current decade (2010-2020): A systematic review. *Open Journal of Modern Linguistics, 11*, 854-872. https://doi.org/10.4236/ojml.2021.116066

Huang, Y. & Gui, M. (2015). 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. https://doi.org/10.11114/jets.v3i3.753

Huddleston, R. & Pullum, G. K. (2002). *The Cambridge grammar of the English language.* Cambridge University Press.

Hung, Y. J. (2019). Bringing assessment and achievement: Repeated practice of self-assessment in college English classes in Taiwan. *Assessment & Evaluation in Higher Education, 44*(8), 1191-1208. https://doi.org/10.1080/02602938.2019.1584783

Idris, M. & Abdul Raof, A. H. (2017). The CEFR rating scale functioning: An empirical study on self- and peer assessments. *Sains Humanika, 9*(4-2). 11-17. https://doi.org/10.11113/sh.v9n4-2.1355

Kang, O. & Yan, X. (2018). Linguistic features distinguishing examinees' speaking performances at different proficiency levels. *Journal of Language Testing & Assessment, 1*, 24-39. https://doi.org/10.23977/langta.2018.11003

Kashinathan, S. & Abdul Aziz, A. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *International Journal of Academic Research in Progressive Education & Development, 10*(2), 983-991. https://dx.doi.org/10.6007/IJARPED/v10-i2/10355

Kaur, M. (2022). Investigation of factors influencing Malaysian ESL learners' English speaking skills**. *International Journal of Advanced Research in Education and Society, 4*(2), 1-6. https://doi.org/10.55057/ijares.2022.4.2.1

Krebs, R., Waldeyer, J., Rothstein, B., & Roelle, J. (2025). Do rubrics enhance self-assessment accuracy? A replication experiment and exploration of effects on performance. *European Journal of Psychology of Education, 40*(1), 1-17. https://doi.org/10.1026/0049-8637/a000296

Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self vs. peer assessment activities in EFL-speaking classes: impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia, 13(*1). https://doi.org/10.1186/s40468-023-00251-3

Li, M. & Zhang, X. (2020). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing, 38*(2), 189-218. https://doi.org/10.1177/0265532220932481

Marzuki, A., Alek, A., Farkhan, M., Deni, R., & Raharjo, A. (2020). Self-assessment in exploring EFL students' speaking skill. *Al-Ta lim Journal, 27*(2), 206-212. https://doi.org/10.15548/jt.v27i2.613
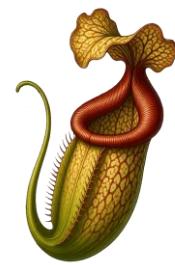
Masruria, W. W. & Anam, S. (2021). Exploring self-assessment of speaking skill by EFL high school students. *Linguistic, English Education and Art, 4*(2), 387-400. https://doi.org/10.31539/leea.v4i2.2285

McMillan, J. H. & Hearn, J, (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1). 40-49. https://eric.ed.gov/?id=EJ815370

Ministry of Education Malaysia (MOE). (2015). Malaysian English language education reform in Malaysia: The roadmap 2015 – 2025. Ministry of Education Malaysia. http://eltc.moe.edu.my/roadmap/The%20Roadmap%202015-2025.pdf

Ministry of Education Malaysia (MOE). (2021). SPM Speaking Examiner Training [PowerPoint slides].

Mohamed Jamrus, M. H. & Razali, A. B. (2019). Using self-assessment as a tool for English language learning. *English Language Teaching. 12*(11). 64-73. https://doi.org/10.5539/elt.v12n11p64

Mohamad Marzaini, A. F., Yusoff, S. M., Zakaria, N., Basal, M. H., Haslee Sharil, W. N. E. & Supramaniam, K. (2024). A discovery on Malaysian ESL teachers' adherence to policy mandates in classroom-based assessment practices. *Malaysian Journal of Learning and Instruction, 21(*1), 249-284. https://doi.org/10.32890/mjli2024.21.1.9

Mohd Don, Z. & Abdullah, M. H. (2019, May 27). What the CEFRis and isn't. Free Malaysia Today. https://www.freemalaysiatoday.com/category/opinion/2019/05/27/what-the-cefr-is-and-isnt/

Mori, Y. & Pell, M.D. (2019). The look of (un)confidence: Visual markers for inferring speaker confidence in speech. *Frontiers in Communication, 4*(63), 1-15. https://doi.org/10.3389/fcomm.2019.00063

Moss, C. M. & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders*. (2nd ed.). ASCD.

Murphy, R. (2019). English grammar in use (5th ed.). Cambridge University Press.

Nadesan, N. K. & Md Shah, P. (2020). Non-linguistic challenges faced by Malaysian students in enhancing speaking skills. *Creative Education, 11*, 1988-2001. https://doi.org/10.4236/ce.2020.1110145

Newlyn, D. (2013). Providing exemplars in the learning environment: The case for and against. *Universal Journal of Educational Research, 1*(1), 26-32. https://doi.org/10.13189/ujer.2013.010104

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychological Review*, *28*, 803-830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E. & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*. https://doi.org/10.1016/j.edurev.2020.100329

Panadero, E., Jonsson, A., Pinedo, L. & Fernandez-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: a meta-analytic review. *Educational Psychology Review*, *35*(4), 113. https://doi.org/10.1007/s10648-023-09823-4

Piccardo, E. & North, B. (2019). *The action-oriented approach: A dynamic vision of language education.* Multilingual Matters.

Rahman, S. A. S. A, Nasarudin, M. R., Rezwan, N. B. S., Zainal Abidin, N. S., Yusri, N. A.. & Arman, N. S. (2025). Investigating public speaking anxiety among undergraduates. *International Journal of Academic Research in Business & Social Sciences, 15*(7), 1499-1520. http://dx.doi.org/10.6007/IJARBSS/v15-i7/26055

Razali, N. H. & Abdul Latif, L. (2019). CEFR-based English speaking skill self-assessment by Malaysian graduating non-native English speaking students. *Malaysian*
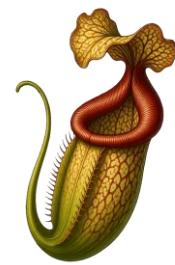
International Journal of Research in Teacher Education, *2*, 82-93. https://mijoriteitepc.wixsite.com/index

Rolheiser, C., & Ross, J. A. (2001). Student self-evaluation: What research says and what practice shows. In R. D. Small, & A. Thomas (Eds.), *Plain talk about kids* (pp. 43-57). Center for Development and Learning.

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*(2), 191–209. https://doi.org/10.1080/0305498870130207

Santos, J. & Ramirez-Avila, M. (2022). Improving speaking fluency through 4/3/2 technique and self-assessment. *Teaching English as a Second or Foreign Language, 26*(2), 1-14. https://doi.org/10.55593/ej.26102a1

Sintayani, N. L. & Adnyayanti, N. L. P. E. (2022). The analysis of self-assessment effect on EFL students' speaking performance. *Journal of Educational Study, 2*(1), 83-91. https://doi.org/10.36663/joes.v2i1.263

Smyth, P. & Carless, D. (2020). Theorising how teachers manage the use of exemplars: towards mediated learning from exemplars. *Assessment Evaluation in Higher Education*, 1-14. https://doi.org/10.1080/02602938.2020.1781785

Su, W. (2020). Exploring how rubric training influences students' assessment and awareness of interpreting. *Language Awareness, 29*(3), 1-19. https://doi.org/10.1080/09658416.2020.1743713

Supramaniam, K., Mohamad Nazer, A. & Maaruf, S. Z. (2020). Teachers' lived-experience in a dual high-stakes examination system. *The European Journal of Social & Behavioral Sciences. 29*(3), 2301-2218. https://doi.org/10.15405/ejsbs.279

To, J., Panadero, E., & Carless, D. (2021). A systematic review of the educational uses and effects of exemplars. *Assessment & Evaluation in Higher Education,* 1-16. https://doi.org/10.1080/02602938.2021.2011134

Ur, P. (2001). *Grammar practice activities*. Cambridge University Press.

Vasileiadou, D. & Karadimitriou, K. (2021). Examining the impact of self-assessment with the use of rubrics on primary school students' performance. *International Journal of Educational Research Open, 2*(2), 1-9. https://doi.org/10.1016/j.ijedro.2021.100031

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Xiao, Y. & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System.* https://doi.org/10.1016/j.system.2019.01.004

Yan, Z., Lao, H., Panadero, E., Fernandez-Castilla, B., Yang, L. & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review, 37*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484

Yufrizal, H. & Pratiwi, N. E. (2020). Self-assessment on communicative competence of students of higher education in Indonesia. *Education Quarterly Reviews, 3*(2), 225-236. https://doi.org/10.31014/aior.1993.03.02.135
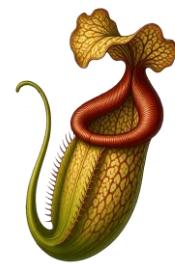
BIOGRAPHY

Vanessa Jacqueline Ligadu is currently pursuing her doctoral degree at the Centre for the Promotion of Knowledge and Language Learning, Universiti Malaysia Sabah. Her areas of interests are ESL teaching and learning, assessment and linguistics.

Priscilla Shak, PhD is a Senior Lecturer at the Centre for the Promotion of Knowledge and Language Learning, Universiti Malaysia Sabah. Her research interests include English for Specific Purposes, technology and AI in language education, language assessment, and curriculum design. She is the main author of Ready to Work: English for Employment (2021) by Cengage Learning and the Social Media Marketing in English Training Module copyright of Universiti Malaysia Sabah.

Appendix A
Speaking Test (Pre-test)

Good morning. First of all, I'd like to know something about you.

Part 1 (3 – 4 minutes)

Candidate A, what is your name?

Candidate B, what is your name?

Candidate A, where do you live?

Candidate B, where do you live?

Now I am going to ask you about your daily routine.

Candidate A, what do you do during your free time?

Candidate B, what do you on weekends?

Part 2 (3 – 4 minutes)

In this part of the test, I'm going to give each of you a topic and I'd like you to talk about it on your own for about a minute. You also need to answer a question briefly about your partner's topic.

Candidate A, it's your turn first. Here's your task. (*Place booklet in front of Candidate A*).

I'd like you to talk about a person you admire. First, you have some time to think about what you're going to say. (*Allow candidate 20 seconds to prepare*).

All right? You may start now.

A person you admire.
Talk about a person you admire. You should say: • Who this person is • What this person looks like • Why you admire this person • If you think it is important for people to have good role models (why / why not?)

Candidate B, who is the person you admire? Why do you admire this person?

Thank you. Can I have the booklet, please?

Now, Candidate B, here's your task. (*Place booklet in front of Candidate B*).

I'd like you to talk about a celebration you had recently. First, you have some time to think about what you're going to say. (*Allow candidate 20 seconds to prepare*).

All right? You may start now.

Thank you. Candidate A, what was the last celebration you had? Did you have a good time? Why?

Can I have the booklet, please?

A celebration you had.

Talk about a celebration you had recently.
You should say:
    • What the event was
    • What you did there
    • If you enjoyed yourself (why / why not?)
Why celebrations are important

Part 3 (4 – 5 minutes)

Now, I'd like you to talk about something together for about three minutes. The task will have two parts. In the first part, you will discuss something with each other for about two minutes. Then, in the second part, you will have another minute to make a decision together. All right? Let's begin with the discussion task.

There are many reasons people use mobile phones. (*Place Part 3 booklet in front of the candidates*).
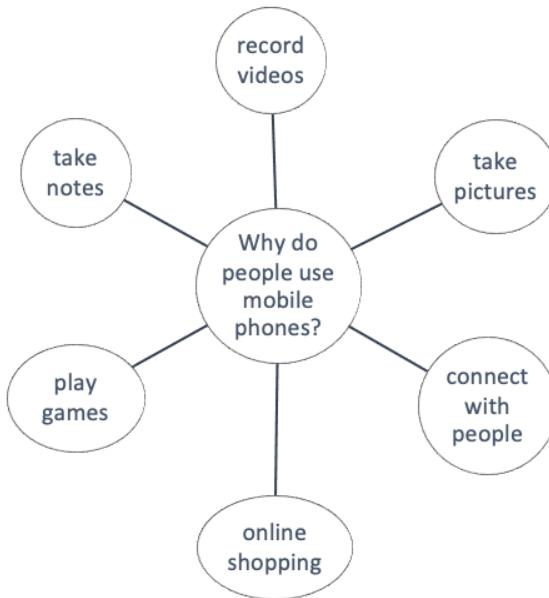
Here are some reasons people use mobile phones and a question for you to discuss. First, have some time to look at the task. (*Allow candidates 20 seconds to prepare*).

Now, talk to each other about the reasons people use mobile phones.

Thank you. Now you have one minute to decide together which is the most popular reason people use mobile phones.

Thank you. Can I have the booklet, please?

You've been talking about the most popular reason people use mobile phones. Now let's hear your opinion on this. What are the advantages and disadvantages of using mobile phones?

Thank you. That's the end of the speaking test.

Appendix B

Speaking Test (Post-test)

Good morning. First of all, I'd like to know something about you.

Part 1 (3 – 4 minutes)

Candidate A, what is your favourite food?

Candidate B, what is your favourite drink?

Candidate A, where do you study?

Candidate B, what is your favourite subject at school?

Now I am going to ask you about your daily routine.

Candidate A, what do you do after school?

Candidate B, what do you do during the school holidays?

Part 2 (3 – 4 minutes)

In this part of the test, I'm going to give each of you a topic and I'd like you to talk about it on your own for about a minute. You also need to answer a question briefly about your partner's topic.

Candidate A, it's your turn first. Here's your task. (*Place booklet in front of Candidate A*).

A teacher you admire.

Talk about a teacher you admire.
 	
You should say:
   • Who this teacher is
   • What this teacher looks like
   • Why you admire this teacher
If you think it is important for teachers to be good role models (why / why not?)

I'd like you to talk about a teacher you admire. First, you have some time to think about what you're going to say. (*Allow candidate 20 seconds to prepare*).

All right? You may start now.

Candidate B, who is the teacher you admire? Why do you admire this teacher?

Thank you. Can I have the booklet, please?

Now, Candidate B, here's your task. (*Place booklet in front of Candidate B*).

I'd like you to talk about what you did during the school holidays. First, you have some time to think about what you're going to say. (*Allow candidate 20 seconds to prepare*).

All right? You may start now.

Thank you. Candidate A, what did you do during the school holiday? Did you have a good time? Why?

Can I have the booklet, please?

Part 3 (4 – 5 minutes)

Now, I'd like you to talk about something together for about three minutes. The task will have two parts. In the first part, you will discuss something with each other for about two minutes. Then, in the second part, you will have another minute to make a decision together. All right? Let's begin with the discussion task.

There are many reasons people use social media. (*Place Part 3 booklet in front of the candidates*).

The school holidays.

Talk about what you did during the school holidays.
You should say:
   • Where did you spend your holidays
   • What you did there
   • If you enjoyed yourself (why / why not?)
Why school holidays are important
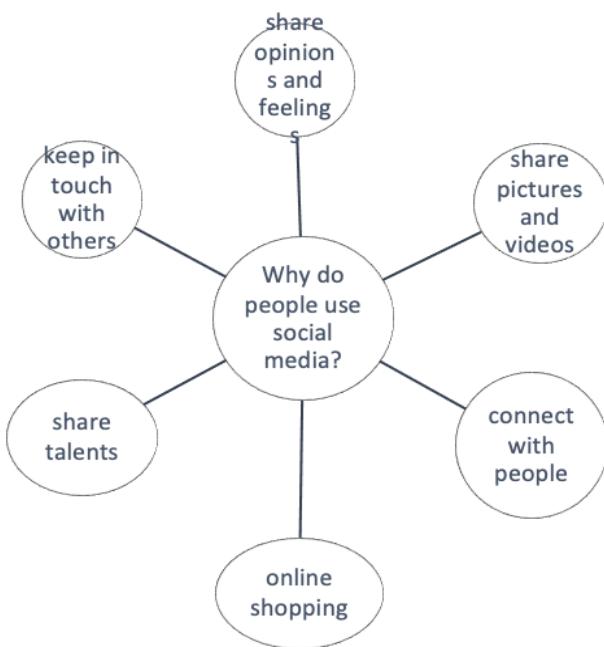
Here are some reasons people use social media and a question for you to discuss. First, have some time to look at the task. (*Allow candidates 20 seconds to prepare*).

Now, talk to each other about the reasons people use social media.

Thank you. Now you have one minute to decide together which is the most popular reason people use social media.

Thank you. Can I have the booklet, please?

You've been talking about the most popular reason people use social media. Now let's hear your opinion on this. Do you think social media has a harmful impact on individuals and society? Why?

Thank you. That's the end of the speaking test.

Appendix C
Speaking Self-Assessment Instrument (SSAI)

Bandscore *Markah band*	1	2	3	4	5	6
1) Grammar *Tatabahasa*	I can talk, but I make many grammar mistakes and sometimes no one can understand me. *Saya boleh bercakap tapi dengan kesalahan tatabahasa yang banyak dan kadangkala, apa yang saya katakan sukar untuk difahami.* I can use only simple structures but with mistakes. *Saya hanya boleh menggunakan struktur ayat yang ringkas tetapi dengan kesalahan.*	I can still make people understand me even if I make grammatical mistakes when I talk. *Saya boleh membuat orang faham apa yang saya katakan walaupun dengan kesalahan tatabahasa.* I can use simple structures with some mistakes, and I sometimes try to use complex structures but with many mistakes. *Saya boleh menggunakan struktur ayat yang ringkas dengan kesalahan tatabahasa dan saya juga boleh menggunakan struktur ayat yang sukar tetapi dengan banyak kesalahan tatabahasa.*	I can talk even if my grammar isn't always right. *Saya boleh bercakap walaupun tatabahasa saya tidak sentiasa betul.* I can use simple structures with a few mistakes, and I try to use some complex structures but with many mistakes. *Saya boleh menggunakan struktur ayat yang ringkas dengan sedikit kesalahan tatabahasa dan saya cuba untuk menggunakan beberapa struktur ayat yang sukar tetapi dengan banyak kesalahan tatabahasa.*	I can speak with almost correct grammar. *Saya boleh bercakap dengan tatabahasa yang hamper betul.* I can use simple and some complex grammatical structures with some errors. *Saya boleh menggunakan struktur tatabahasa yang mudah dan kompleks dengan beberapa kesalahan.*	I can talk with correct grammar most of the time. *Saya boleh bercakap dengan tatabahasa yang betul pada kebanyakan masa.* I can use simple and some complex grammatical structures with very few errors. *Saya boleh menggunakan struktur tatabahasa yang mudah dan kompleks dengan kesalahan yang sangat sedikit.*	I can use correct grammar when I talk. I can correct my own grammar mistakes on the spot. *Saya boleh menggunakan tatabahasa yang betul semasa bercakap. Saya boleh membetulkan kesilapan tatabahasa saya.* I can use a wide range of grammatical structures well. *Saya boleh menggunakan pelbagai struktur tatabahasa dengan baik.*

Bandscore / Markah band	1	2	3	4	5	6
2) Vocabulary / *Kosa kata*	I can use sufficient vocabulary to talk about everyday situations and familiar topics. *Saya boleh menggunakan perbendaharaan kata yang mencukupi untuk bercakap tentang situasi harian dan topik biasa.* I mostly can convey meaning but may use inappropriate vocabulary. *Saya kebanyakannya boleh menyampaikan maksud tetapi mungkin menggunakan perbendaharaan kata yang tidak sesuai.*	I can use sufficient vocabulary to talk about everyday situation and familiar topics. *Saya boleh menggunakan perbendaharaan kata yang mencukupi untuk bercakap tentang setiap situasi harian dan topik biasa.* I can convey relevant meaning but may always have errors in vocabulary choice. *Saya boleh menyampaikan maksud yang relevan tetapi mungkin sentiasa mempunyai kesilapan dalam pilihan perbendaharaan kata.*	I can use appropriate vocabulary to talk about everyday situations and familiar topics. *Saya boleh menggunakan perbendaharaan kata yang sesuai untuk bercakap tentang situasi harian dan topik biasa.* I can convey relevant meaning but may have errors in vocabulary choice. *Saya boleh menyampaikan maksud yang relevan tetapi mungkin mempunyai kesilapan dalam pemilihan kosa kata.*	I can use a range of appropriate vocabulary to give and exchange views on familiar topics. *Saya boleh menggunakan pelbagai perbendaharaan kata yang sesuai untuk memberi dan bertukar-tukar pandangan tentang topik biasa.* I can convey relevant meaning with sometime good vocabulary choice. *Saya boleh menyampaikan maksud yang relevan dengan pilihan kosa kata yang baik.*	I can use a wide range of appropriate vocabulary to give and exchange views on a wide range of familiar topics. *Saya boleh menggunakan pelbagai perbendaharaan kata yang sesuai untuk memberi dan bertukar-tukar pandangan tentang pelbagai topik biasa.* I can convey relevant meaning with good vocabulary choice. *Saya boleh menyampaikan maksud yang relevan dengan pemilihan kosa kata yang baik.*	I can use a wide range of appropriate vocabulary to give and exchange views on a wide range of abstract, complex and unfamiliar topics. *Saya boleh menggunakan pelbagai perbendaharaan kata yang sesuai untuk memberi dan bertukar pandangan tentang pelbagai topik yang abstrak, kompleks dan asing.* I can convey specific differences in meaning through appropriate choice of vocabulary. *Saya boleh menyampaikan perbezaan makna yang khusus melalui pemilihan kosa kata yang sesuai.*
Bandscore / Markah band	1	2	3	4	5	6
3) Communicative Competence	I have difficulties to start, maintain and end a conversation.	I have some difficulties to start, maintain and end a conversation.	I can start a conversation, maintain,	I can start a conversation, maintain, develop and end it.	I can easily start a conversation, maintain, develop and end it.	I can excellently start a conversation, maintain, develop and end it.

Kecekapan berkomunikasi	*Saya menghadapi kesukaran untuk memulakan, mengekalkan dan menamatkan perbualan.* I have difficulties to ask for clarification when necessary. *Saya menghadapi kesukaran untuk meminta penjelasan apabila perlu.*	*Saya mempunyai beberapa kesukaran untuk memulakan, mengekalkan dan menamatkan perbualan.* I have some difficulties to ask for clarification when necessary. *Saya mempunyai sedikit kesukaran untuk meminta penjelasan apabila perlu.*	develop and end it with little difficulty. *Saya boleh memulakan perbualan, mengekalkan, mengembangkan dan menamatkannya dengan sedikit kesukaran.* I have little difficulties to ask for clarification and further details to move a discussion forward. *Saya mempunyai sedikit kesukaran untuk meminta penjelasan dan butiran lanjut untuk menggerakkan perbincangan ke hadapan.*	*Saya boleh memulakan perbualan, mengekalkan, mengembangkan dan menamatkannya.* I can ask for clarification and further details to move a discussion forward. *Saya boleh meminta penjelasan dan butiran lanjut untuk menggerakkan perbincangan ke hadapan.*	*Saya boleh dengan mudah memulakan perbualan, mengekalkan, mengembangkan dan menamatkannya.* I can work well and negotiate towards an outcome with my partner. *Saya boleh bekerjasama dengan baik dan berunding untuk mendapatkan keputusan dengan pasangan saya.*	*Saya boleh memulakan perbualan, mengekalkan, mengembangkan dan menamatkannya dengan cemerlang.* I can work well and effectively with my partner towards an outcome. *Saya boleh bekerjasama dengan baik dan berkesan dengan pasangan saya ke arah mendapatkan persetujuan bersama.*
Bandscore *Markah band*	**1**	**2**	**3**	**4**	**5**	**6**
4) Fluency *Kelancaran*	I can make myself understood but may have a lot of pauses, hesitations and false starts. *Saya boleh membuat diri saya difahami tetapi mungkin*	I can make myself understood but may have a lot of pauses and hesitation. *Saya boleh membuat diri saya difahami tetapi mungkin banyak terdiam dan teragak-agak.*	I can express myself with some pauses and hesitation. *Saya boleh menjelaskan diri saya dan sekali-sekala terdiam dan teragak-agak.*	I can express myself sometimes with some or little pauses and hesitation. *Saya boleh menjelaskan diri saya kadang-kadang dengan sedikit teragak-agak.*	I can express myself with little pauses and hesitation. *Saya boleh menjelaskan diri saya dengan sedikit teragak-agak.*	I can express myself spontaneously without pauses and hesitation. *Saya boleh mengekspresikan diri saya secara spontan tanpa terdiam dan teragak-agak.*

	banyak terdiam, teragak-agak dan permulaan yang salah.					
Bandscore *Markah band*	1	2	3	4	5	6
5) Pronunciation *Sebutan*	I can speak with unclear pronunciations all the time *Saya boleh bercakap dengan sebutan yang tidak jelas sepanjang masa.*	I can speak with unclear pronunciation which makes it difficult to understand me. *Saya boleh bercakap dengan sebutan yang tidak jelas yang menyukarkan untuk saya difahami.*	I can sometimes speak clearly and mispronounce many words. *Saya kadang-kadang boleh bercakap dengan jelas dan menyebut banyak perkataan dengan salah.*	I can speak clearly most of the time and mispronounce three to five words. *Saya boleh bercakap dengan jelas pada kebanyakan dan salah menyebut tiga hingga lima perkataan.*	I can speak clearly all the time and mispronounce one or two words. *Saya boleh bercakap dengan jelas sepanjang masa dan salah sebut satu atau dua perkataan.*	I can speak clearly all the time and not mispronounce words. *Saya boleh bercakap dengan jelas sepanjang masa dan tidak salah menyebut perkataan.*
Bandscore *Markah band*	1	2	3	4	5	6
6) Topic Management *Pengurusan topik*	I can sometimes show understand of one of the topics discussed. *Saya kadangkala boleh menunjukkan kefahaman tentang salah satu topik yang dibincangkan.*	I can show understanding of few of the topics discussed. *Saya boleh menunjukkan pemahaman tentang beberapa topik yang dibincangkan.*	I can show good understanding of some of the topics discussed. *Saya dapat menunjukkan pemahaman yang baik tentang beberapa topik yang dibincangkan.*	I can show good understanding of most of the topics discussed. *Saya boleh menunjukkan pemahaman yang baik tentang kebanyakan topik yang dibincangkan.*	I can show a good understanding of almost all of the topics discussed. *Saya dapat menunjukkan pemahaman yang baik tentang hampir semua topik yang dibincangkan.*	I can show full understanding of all of the topics discussed. *Saya boleh menunjukkan pemahaman penuh tentang semua topik yang dibincangkan.*

Bandscore *Markah band*	1	2	3	4	5	6
7) Confidence *Keyakinan*	I can always talk about the topics in an uncertain way. *Saya sentiasa boleh bercakap tentang topik dengan cara yang tidak pasti.*	I can talk about the topics in an uncertain way most of the time. *Saya boleh bercakap tentang topik dengan cara yang tidak pasti pada kebanyakan masa.*	I can talk about the topics in an uncertain way sometimes. *Saya boleh bercakap tentang topik dengan cara yang tidak pasti kadang-kadang.* I seldom use eye contact, facial expressions and body movements as I speak. *Saya jarang menggunakan pandangan mata, mimik muka dan pergerakan badan semasa bercakap.*	I can talk about the topics in a confident way sometimes. *Saya boleh bercakap tentang topik dengan cara yang yakin kadang-kadang.* I sometimes use eye contact, facial expressions and body movements as I speak. *Saya kadang-kadang menggunakan pandangan mata, mimik muka dan pergerakan badan semasa bercakap.*	I can talk about the topics in a confident way most of the time. *Saya boleh bercakap tentang topik dengan cara yang yakin pada kebanyakan masa.* I frequently use eye contact, facial expressions and body movements as I speak. *Saya kerap menggunakan pandangan mata, mimik muka dan pergerakan badan semasa bercakap.*	I can talk about the topics in a confident way all the time. *Saya boleh bercakap tentang topik dengan cara yang yakin sepanjang masa.* I always use eye contact, facial expressions and body movements as I speak. *Saya selalu menggunakan pandangan mata, mimik muka dan pergerakan badan semasa bercakap.*

Bandscore *Markah band*	1	2	3	4	5	6
8) Organisation *Organisasi*	I can use connectors to link simple sentences. *Saya boleh menggunakan kata hubung untuk menghubungkan ayat mudah.*	I can sometimes form longer sentences and may have errors linking them together using basic cohesive devices. *Saya kadangkala boleh membuat ayat yang lebih panjang dan mungkin mempunyai kesilapan menghubungkannya*	I can form longer sentences and link them together using basic cohesive devices. *Saya boleh membuat ayat yang lebih panjang dan menghubungkannya menggunakan penanda wacana asas.*	I can sometimes use a range of cohesive devices but may have errors. *Saya kadangkala boleh menggunakan pelbagai penanda wacana tetapi mungkin mempunyai kesalahan.*	I can use a range of cohesive devices. *Saya boleh menggunakan pelbagai penanda wacana.*	I can use a wide range of organisational patterns, cohesive devices and connectors. *Saya boleh menggunakan pelbagai jenis organisasi, penanda wacana dan kata hubung.*

				menggunakan penanda wacana asas.		
Bandscore *Markah band*	**1**	**2**	**3**	**4**	**5**	**6**
9) Strategy Use *Penggunaan strategi*	I avoid unfamiliar language all the time. *Saya mengelak apabila tidak faham sepanjang masa.*	I avoid unfamiliar language most of the time. *Saya mengelak apabila tidak faham kebanyakan masa.*	I avoid unfamiliar language sometimes. *Saya mengelak apabila tidak faham kadang-kadang.*	I can deal with unfamiliar language. *Saya boleh mengatasi bahasa yang saya tidak faham.* I compensate by using familiar language sometimes. *Saya menggantikan bahasa yang sukar dengan menggunakan bahasa yang lebih senang.*	I can deal with unfamiliar language. *Saya boleh mengatasi bahasa yang saya tidak faham.* I know most of the language used in the tasks. *Saya tahu kebanyakan bahasa yang digunakan dalam tugasan.*	I can deal with unfamiliar language. *Saya boleh mengatasi bahasa yang saya tidak faham.* I know all of the language use in the tasks. *Saya tahu semua penggunaan bahasa dalam tugasan.*
Bandscore *Markah band*	**1**	**2**	**3**	**4**	**5**	**6**
10)Time Management *Pengurusan masa*	I can speak only for a limited time. *Saya boleh bercakap hanya untuk masa yang terhad.*	I can sometimes speak for more than 1 minute. *Saya kadang-kadang boleh bercakap selama lebih daripada 1 minit.*	I can sometimes speak within the allotted time given. *Saya kadang-kadang boleh bercakap dalam masa yang ditetapkan.*	I can speak within the time given with ease. *Saya boleh bercakap dalam masa yang diberikan dengan mudah.*	I can speak within the time given well. *Saya boleh bercakap dalam masa yang diberikan dengan baik.*	I can speak within the time given excellently well. *Saya boleh bercakap dalam masa yang diberikan dengan baik.*

Adapted from the SPM English – Speaking assessment scales (Ministry of Education, 2021), CEFR Can-do statements (Council of Europe, 2020; Piccardo & North, 2019), Babaii *et al.* (2015), & Mori & Pell (2019).