Dear Second Language Learners (L2): The Complete Guide of Primary Trait Scoring Rubric for Self and Peer Assessment (SAPA)

Nurul Fariena Asli, Mohd Effendi Ewan Mohd Matore

In-Text Citation: (Asli & Matore, 2023)
To Cite this Article: Asli, N. F., & Matore, M. E. E. M. (2023). Dear Second Language Learners (L2): The Complete Guide of Primary Trait Scoring Rubric for Self and Peer Assessment (SAPA). *International Journal of Academic Research in Progressive Education and Development*, *12*(2), 2678–2693.

Full Terms & Conditions of access and use can be found at
http://hrmars.com/index.php/pages/detail/publication-ethics

Dear Second Language Learners (L2): The Complete Guide of Primary Trait Scoring Rubric for Self and Peer Assessment (SAPA)

Nurul Fariena Asli, Mohd Effendi Ewan Mohd Matore

Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Malaysia
Corresponding Author's Email: effendi@ukm.edu.my

Abstract
In assessing performance-based language assessment, the use of a suitable scoring method is crucial to minimize measurement errors that will become threats in the rating process. The scoring method which is widely used in the literature is a rubric as the rubric is proven to be able to provide guidelines for the rater to cognitively construct their understanding of how to assess performance. The appropriate selection of a rubric to be used by a particular rater will ensure the validity and reliability of the scores given to an assessed performance. However, since there are different types of rubrics and each of the rubrics serves a different purpose, teachers must be able to distinguish the differences between these rubrics. Therefore, this paper aims to discuss the different types of rubrics and highlight the potential of primary trait rubrics to be used by second language learners (L2) in one of the performance-based assessment activities particularly self and peer assessment (SAPA). The main idea of this paper is to improve the quality of formative assessment practice in measuring second language performance which is significant to ensure the validity and reliability of the assessed performance. Undeniably, this will provide insights for teachers to conduct the formative assessment in the classroom mainly when self-directed learning is becoming the main aim for students' development.
Keywords: Primary Trait Rubric, Holistic Rubric, Analytic Rubric, Performance Assessment, Classroom Assessment, Self-Assessment, Peer Assessment

Introduction

Major changes have been made to the assessment of second language (L2) learners' English proficiency and these have led countries, particularly in Southeast Asia, to integrate the Common European Framework of Reference (CEFR) into the education system, including Malaysia. In 2017, the Malaysian Ministry of Education (MOE) introduced the CEFR to improve the English language proficiency of Malaysian students and to set a benchmark for international standards (Kaur & Zhi, 2022). The alignment of the curriculum with CEFR is one of the efforts by MOE to shift the exam-oriented system which has been the main practice in Malaysia towards formative assessment. This is since the previous written curriculum is not always brought life to the classroom and examinations are failed to assess the full range of

skills that the education system aims to produce (Don, 2015). The CEFR has provided new insights into how language assessment should be conducted as it emphasizes performance-based language assessment compared to the traditional method of assessing language proficiency.

Mcnamara (1996) distinguishes between traditional assessment and performance assessment by adding two important elements, the rater and the rating scale (rubric). As a tool that can help raters make accurate and reliable judgments in an evaluation process, the use of rubrics has increased in both research and practice (Dickinson & Adams, 2017). Rubrics are also recognized because they help clarify teachers' expectations, highlight students' strengths and weaknesses, and encourage students to self-assess (Becker, 2016). However, when using rubrics, some raters are highly susceptible to subjective judgments where there can be inconsistencies and a tendency to be lenient or harsh (Trace et al., 2016) This makes it difficult to ensure that they are fulfilling their obligation to provide meaningful assessments (Martens, 2018) and to evaluate the quality and success of performance beyond a simple description (Dickinson & Adams, 2017).

In terms of self and peer assessment (SAPA), the Roadmap for the English Language in Malaysia has highlighted that this is one of the activities that needs to be encouraged in the classroom as it helps to empower students to be responsible for their own learning (Mohd Don, 2015). SAPA is the culmination of the formative assessment context as it encourages students to be independent learners. To ensure the implementation of SAPA in the classroom, an assessment method that is an effective rubric is needed to support students, especially second language (L2) learners. This is evidenced by the fact that most SAPA studies have used rubrics to test students' language proficiency (Idris & Abdul Raof, 2017), assess the validity of a score (Deygers & Van Gorp, 2015), identify assessment errors (Ahmadi Shirazi, 2019; Engelhard & Wind, 2019), and determine the effectiveness of teacher and learner assessment activities (Mohavedi & Kiasi, 2021). The purpose of this paper is to illustrate the potential of the primary trait rubric, which can be used in SAPA activities to assess students' performance. In addition, the advantages and disadvantages of other rubrics will also be discussed to verify why the primary trait rubric is best suited for SAPA.

Why is a suitable rubric needed to be used in SAPA?
With the incorporation of the CEFR into the curriculum, teachers in Malaysia will become familiar with the CEFR grading scales and will be required to align their teaching and assessment with the scales. The CEFR contains grading scales that provide curriculum developers, lecturers, and teachers with a set of benchmarks to which they can refer. However, there are problems with the CEFR scales as it is challenging for language assessors to use them due to the lack of consistency, validity, and applicability of the existing descriptors (Simons & Colpaert, 2015; Zou & Zhang, 2017). On the other hand, North (2014) argues that the high degree of generality of the scales allows them to be contextualized for a variety of situations and does not limit the user's ability to modify and adapt student learning. Moreover, the CEFR scales are designed to measure competencies only and have a different purpose than assessment scales (Holzknecht et al., 2018). Therefore, they are not intended to be used exclusively for rating purposes.

Thus, in a study by Lee & Kassim (2019), it was found that 25 out of 30 teachers found it difficult to develop assessments based on the CEFR because of their poor understanding of the assessment criteria. This has influenced teachers' controversial decision to incorporate assessments and lessons linked to the CEFR into their teaching (Lee & Kassim, 2019). This is also due to the fact that there is no assessment system or strategy available to teachers especially when conducting lesson-based assessments (Veloo et al., 2018). Because grades are assigned based on rater preferences rather than a specific standard, it would be possible to debate the grades assigned to students in the absence of a grading rubric. Raters may have different views of what constitutes good performance standards. In most cases, this would lead to measurement inaccuracies among raters, and therefore the grades assigned would not be valid or reliable because they would not reflect actual student performance.

The 2015-2025 English Language Roadmap has highlighted the need to promote students who are aware of their weaknesses and strengths to enable independent learning (Mohd Don, 2015). In line with this, the current CEFR curriculum aligned to the KSSM has been underpinned by the principle of Assessment as Learning (AaL), such that students need to be trained and empowered to be assessors who can make decisions for their learning (Kayapinar, 2014). While many teachers use rubrics to measure language learning, learners are rarely involved in their use and development as raters. It is well known that before the current reforms, the Malaysian education system was examination-based and relied solely on teachers as the primary assessors in all assessments in the Malaysian classroom. Even with the introduction of school-based assessment (SBA) in 2015, students are still assessed by teachers. This has led to a situation where students are rarely asked to evaluate themselves or their classmates, as the system is skeptical of trusting students with the position of evaluator (Idris & Abdul Raof, 2017).

Since entrusting students with the assessment of their learning is not without its problems, having a valid and reliable rubric that they can use is crucial to reduce the measurement errors that might be caused by them in an assessment process. Since students in Malaysia can be considered L2 learners, they need to use a different assessment method because L2 learners need a very specific guide to help them (Weigle, 2010). In general, there are a few types of rubrics used to assess student performance, namely holistic, analytic, and primary trait rubrics. Holistic and analytic rubrics are most commonly used, while the primary trait rubric is not widely used in research, although it is claimed that it is more suitable for L2 learners due to its complexity in creation (Othman, 2014; Veloo et al., 2018; Weigle, 2010). Based on the characteristics of these three rubrics, which will be discussed in detail in the next section, it is hypothesized that the primary trait rubric has the greatest potential to be used by L2 learners at SAPA. Before addressing all rubrics, it is necessary to review the changes that were made in the assessment system, especially when the CEFR was aligned with the English language curriculum. The context for this document is the secondary school curriculum.

Literature Review
Formative Assessment in KSSM-aligned CEFR

In March 2018, a teacher's guide entitled Assessment Principles and Practices" for Secondary Teachers was published as one of the modules teachers must cover during preparation for the CEFR course. The manual was produced by Cambridge Assessment English

in collaboration with the Malaysian Ministry of Education (MOE) to promote the use of formative assessment in the classroom. It is an introduction to the concepts and methods of formative assessment, presenting teachers with practical recommendations and ideas for classroom use. In the current curriculum, formative assessment is supposed to be at the center of teaching and learning because formative assessment has a significant positive effect on learning (Black & Wiliam, 2009). It has been emphasized that formative assessment is the bridge between teaching and learning in every lesson, with research showing that it can accelerate student learning by building this bridge.

Formative assessment should not be considered separately from teaching and learning because they are components of an ongoing process in which there are three main questions that every teacher must consider in his or her teaching and learning process: "Where are we going?"," "Where are the students now?"," and "How do we get there?" These questions suggest that learning is a system that requires constant communication between teachers and students. The elements of the teacher-student relationship in formative assessment are to ensure that each child understands what he or she is learning and how to measure his or her own success. In addition, formative assessment can allow students to create data that reflects their learning and allow teachers and students to track and coordinate their learning.

It is also a must for educators to consistently provide feedback to students so they can improve their learning. This will help improve individual student language acquisition, and of course, it will improve understanding of student needs when implementing formative assessment practices. There are two other important aspects of formative assessment: how can students support each other and how can they become independent learners and engage in their own learning. For this reason, self and peer assessment (SAPA) is very important to formative assessment as we incorporate it into independent learning.

In formative assessment, Black and Wiliam (2018) suggested three different periods for responses, and each cycle has a different type of effect. The period is within and between classes for short cycles, and it is within and between classes for medium cycles, but it is by semester or class for long cycles. The duration of the short cycle is minute to minute and day to day, and the result is that it would improve student interaction and teacher responsiveness. The medium cycle lasts one to four weeks and has the effect of helping students understand how they are being measured and helping teachers develop a better understanding of assessment. Finally, the long cycle lasts four weeks to a year and has the effect of measuring student success and making changes to the curriculum (Black & Wiliam, 2018).

Table 1

Formative Assessment Cycle

	Short cycle	Medium cycle	Long cycle
Span	Within and between lessons	Within and between teaching units	Across terms or teaching units
Length	Minute-by-minute and day-by-day	One to four weeks	Four weeks to a year (or more)
Impact	Pupil engagement increases Teacher responsiveness improves	Pupils understand how they are assessed Teacher understanding of assessment improves	Benchmarking for pupil progress Improvements made to the curriculum

Self-Assessment and Peer Assessment

The formative assessment SAPA is discussed further in this section as it provides the platform for participants in this study to complete the written assessment. As Malaysia is still in the second phase of the English language roadmap, it is still the phase to practice the integration of assessment practices that promote greater student independence and self-directed learning (Don, 2015). Therefore, it is assumed that students are not yet familiar with SA or have not been exposed to the activity at this stage. This is based on the findings of recent studies on SBA, where there was limited evidence of SAPA in schools (Idris & Abdul Raof, 2017; Sidhu et al., 2018). In this section, SA and PA are discussed separately to provide a clearer understanding of each of the practices and are later combined for the needs of this study.

To become a successful autonomous learner, as the new KSSM curriculum aims to do, the ability to monitor one's actions, thoughts, and feelings to achieve set goals is crucial (Panadero & Romero, 2014). For this reason, SA is one of the formative assessments that teachers should conduct in the classroom (Don, 2015), as it can improve students' self-regulation in learning. SA can be defined as a process of formative assessment in which students reflect on and assess their work and learning based on explicitly shared goals or success criteria, and then identify strengths and weaknesses and make revisions to improve them (Andrade & Du, 2007). A simpler definition by Andrade (2019) for SA refers to all assessments that learners make about their work.

Andrade and Du (2007) also proposed a three-step pedagogical process in which the three steps are (1) the teacher explains the intended performance expectations to students, (2) students complete their assignment and review their work using the rubric provided, and (3) students revise and improve their work based on feedback received at SA. This is similar to the steps suggested by (Brown et al., 2015). It is suggested that the three underlying principles and steps in SA are: (1) formulation of expectations, (2) self-assessment, (3) revision. Thus, the critical step in conducting SA is for the teacher to define the expectations for a particular task and explain to students each criterion used in SA. A thorough discussion with students about the criteria of a rubric in a written assessment activity is necessary so that students know what is expected of them. In the second step, students write their essays

according to the assigned task and practice self-assessment using the criteria discussed earlier. They can note down the mistakes they made when they did not achieve the intended goals. In the last step, students work out the final assignment and compare it with their teacher.

Peer assessment (PA) or sometimes known as peer evaluation among students, is considered one of the best grading methods that requires student input (Rahadi et al., 2018). Reinholz (2016) defined PA as a set of activities in which individuals make judgments about the work of others. In the classroom, students are required to provide feedback or evaluation to their peers on a performance task based on a rubric or criteria that students have engaged with (Mumpuni et al., 2022). Here, students can practice helping each other and are not solely dependent on teachers to evaluate and make judgments about their work.

The benefits of PA are multiple: (1) learning gaps can be better understood when students help each other rather than receiving it from the teacher, as this improves students' status in learning (Mumpuni et al., 2022), (2) it promotes a better understanding of the criteria needed to achieve the performance goal, as students are exposed to different types of responses from peers (Reinholz, 2016), (3) they develop critical reflection skills based on others' responses (Wanner & Palmer, 2018). Schunn et al (2016) have identified two features that are critical for PA, namely that students must be able to provide sincere feedback and that teachers' fear of the social implications of honest peer review must be reduced. If these two issues are not resolved, the validity and reliability of PA will be compromised. This is evident in a study by Schunn et al (2016) on the validity and reliability of PA in a secondary school writing assessment. The study was conducted due to the lack of focus on the validity and reliability of PA and the result is that it can be minimized if a good rubric is created for PA as well as SA. Therefore, the task of this study is to investigate SAPA using the most appropriate rubric after reviewing the literature.

Types of Scoring Rubrics

It is a very delicate task to test a written assessment, and the research landscape in recent decades has focused on the use of the assessment rubric as one of the tools of the criterion-based methodology. Based on the parameters related to the assessment rubric, the content of each essay is assessed, e.g., coherence, grammatical consistency, contextual appropriateness, and so on. A rubric can be described as a set of parameters containing details of the performance standard (Brookhart & Chen, 2015) that is explicitly created for an assessment to be administered by those who use it. It is also the criterion against which an assessment is evaluated (Francis, 2018) and is intentionally used to achieve such goals for students. Rating scales are different, as a rubric does not provide definitions of the level of performance in addition to the parameters. Rating scales tend to be widely used and can be interpreted depending on the context. Rubrics are derived from rating scales but become more specific depending on the intended function of their creators.
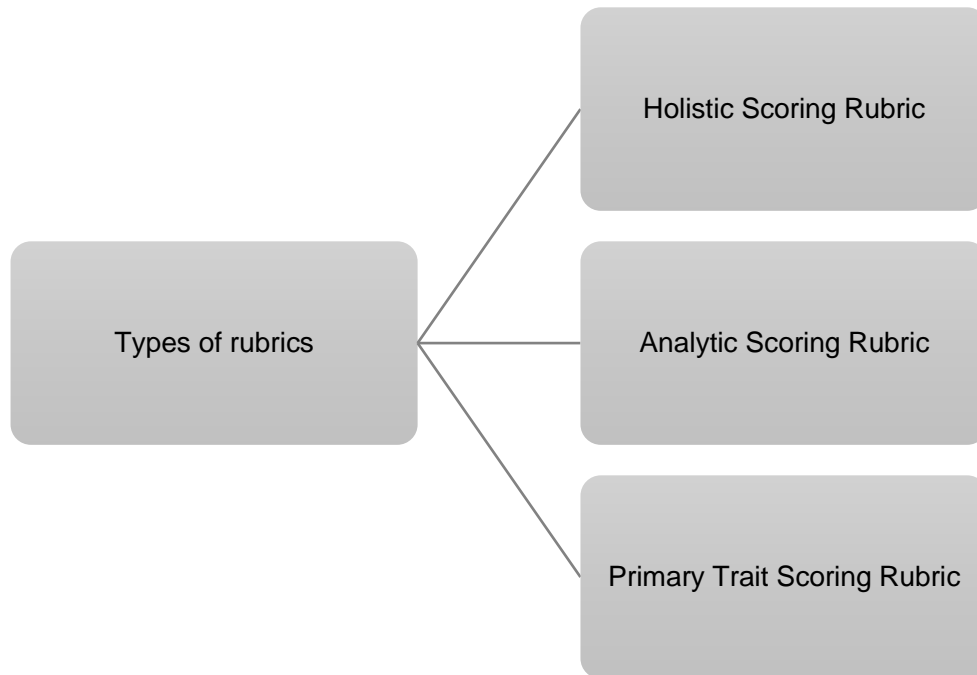
```
                                          ┌─────────────────────────┐
                                          │                         │
                                          │ Holistic Scoring Rubric │
                                          │                         │
                                          └─────────────────────────┘
  ┌─────────────────────┐                 ┌─────────────────────────┐
  │                     │                 │                         │
  │  Types of rubrics   │────────────────│ Analytic Scoring Rubric │
  │                     │                 │                         │
  └─────────────────────┘                 └─────────────────────────┘
                                          ┌─────────────────────────────┐
                                          │                             │
                                          │ Primary Trait Scoring Rubric│
                                          │                             │
                                          └─────────────────────────────┘
```

Figure 1: Types of rubrics

Holistic Scoring Rubric

The holistic assessment rubric is one of the most common forms of rubrics used in performance assessment. In this scoring method, a single grade is given to the examinee on the consistency of the examinee's work as a whole. Thus, when scoring an essay holistically, the scorer considers the entire content of the paper and assigns a grade. This allows for a quick assessment of the results of an essay. This saves time, especially when a large number of essays need to be evaluated at the same time. The holistic assessment approach is commonly used to score an essay and is ideal for large-scale assessment because it is time and cost efficient (Veloo et al., 2018). An intriguing example of holistic scoring is that for poorly structured essays with few grammatical errors, a well-organized essay with several grammatical errors can receive the same score. This is due to the fact that the rater will first analyzes all the strengths and weaknesses of the essay based on the given criteria and then tries to assign the best grade. Raters are expected to consider many different features and provide an overall impression with a score.

Therefore, in addition to sample essays, it is necessary to train raters with a holistic rubric to prepare raters for holistic marking. The sample essays would provide an example of the performance levels in each rubric area. Using this sample essay, raters can see how to distinguish the characteristics of a high-scoring essay and a low-scoring essay. If raters choose to score the essays themselves, they will refer to the best match in the band for the sample essays. Compared to other forms of rubrics, less and shorter time is required to train raters to use a holistic rubric (Veloo et al., 2018).

The use of a holistic rubric raises several issues that are considered to challenge the reliability and validity of an assessment. This is because holistic descriptors lead examiners to think of a number of different linguistic elements when analyzing the student's essay, but in essence, the different aspects of success feed into a single score (Davis, 2018). Examiners may

encounter difficulties and uncertainties in assessing and awarding the correct marks for the essay. The essay does not provide information about students' strengths and deficiencies of students in terms of classroom assessment. As Weigle (2010) has shown, it could not make a detailed contribution to the fact that the holistic approach to grading focuses on the overall impression of the writing. It only focuses on what writers are "good at" based on their general characteristics rather than finding their incompetence and weaknesses in writing (Salmani Nodoushan, 2014). Ohta et al (2018) have emphasized that in L2 writing instruction, the small amount of knowledge accessible to test users through the use of the holistic scoring rubric is problematic.

Although there are many drawbacks to using the holistic rubric in grading, the findings of Veloo et al (2018) showed that the majority of ESL teachers in their study preferred the holistic assessment method because of its time-saving feature when extensive grading is required, and it is important to complete the results within a short period. The study also showed that teachers believe that the holistic assessment approach enables them to improve their students' learning because they can measure their students' overall success. This finding contrasts with Hamp-Lyons' (1995) argument that second-language examiners find the holistic assessment approach problematic. Namely, she felt that some ESL examiners seem to focus on one particular aspect when assessing their students' written work.

For example, some examiners pay more attention to students' grammatical errors than to their ability to elaborate on points and to the accuracy of the essay. Ohta et al (2018) confirmed this argument, saying that their context, such as cultural, linguistic, and educational factors, might contribute to their judgment in evaluating students' essays when raters use a holistic scale. In addition, Nodoushan (2014) found that ESL raters appear to be influenced by the length of the essay and the need to avoid making basic technical errors when rating their students' essays. Thus, she suggests that teachers prefer the holistic scoring rubric when evaluating their students' essays, but teachers must not disregard the danger involved.

Based on the above clarification, it may not be sufficient to use a holistic form of scoring rubric for small-scale evaluation purposes, such as in classroom-based assessment activities. In addition, expert raters tend to be more accurate in the use of a holistic rubric, as an expert decision is needed to evaluate an essay. As this study will use students as raters, it is not acceptable to use a holistic rubric in this study.

Analytic Scoring Rubric

An analytic scoring rubric is a tool used to assess the work of students, which involves assigning individual scores for different parts of an assignment. It is increasingly used for assessment of performance to indicate the various aspects of a task so that students can be made aware of their performances on each aspect by reference to an appropriate set of level requirements. To check various elements or aspects of the output on a corresponding score and parameter, examiners and raters are using an analytical rubric (Idris & Abdul Raof, 2017). It is important to define the language features that "count" and also to consider the relative weight given to each feature for analytic scales (Davis, 2018). Analytic scoring is used in a variety of settings, particularly in education, to assess student achievement, e.g., language arts, science, mathematics, art, etc. It is also used to test students in engineering, nursing,

business, and teaching who are preparing for careers. It is also designed for all grade levels, from preschool through university. Analytic scoring allows raters to assess the scores of the written product by evaluating the various aspects or characteristics of the writing. The scoring procedure uses separate scales, with each scale measuring a specific element of writing. Most of the subdivided features in the writing assessment are the development of ideas, organization, language use, vocabulary, grammar, spelling, and writing mechanics (Idris & Abdul Raof, 2017; Othman, 2014).

There are many advantages of the analytical scoring rubric that have led researchers to focus on this assessment rubric in their studies. First and foremost, the use of an analytic rubric has been shown to enhance learning by making expectations and criteria clear. Ghalib and Al-Hattami (2015) state that analytic scoring allows for the separation of the various features of a composition into components for assessment. It is very precise and comprehensive as every linguistic aspect is considered in the evaluation of the performance. In addition, the analytic scoring rubric provides teachers with comprehensive feedback and helps them evaluate the weak and strong aspects of students' written performance. It has a higher discriminatory power that provides information about students' abilities (Ghalib & Al-Hattami, 2015) and justifies uneven performance with very detailed feedback to examinees (Davis, 2018).

However, the results of the study Veloo et al (2018) suggest that the analytic scoring rubric takes time and that raters need to refer back to the rubrics from time to time when evaluating student writing because of the characteristics explained in the previous paragraphs. Davis (2018) supports this because the analytic rubric allows raters to make multiple scoring decisions at a given time when reviewing a student's writing. In Malaysia, this form of assessment is ideal for a single class with fewer students, while the majority of students in a classroom are 35-40 students. Because every element of language is considered and assessed, the analytical assessment method is detailed. It is time-consuming but provides comprehensive information to distinguish which aspects of students' (writers') writing are strong or weak. For small-scale assessments with a smaller number of students, the analytic assessment approach is sufficient (Veloo et al., 2018).

Primary Trait Scoring Rubric
The primary trait scoring is another rating scale that is considered very useful for evaluating written work, especially for L2 learners. It was first introduced and developed in the early 1970s by Lloyd Jones for a large-scale school testing programme under the National Assessment of Educational Progress (NAEP) in the United States. Since then, it has been used by researchers in a variety of ways, including the assessment of task-based second language tests. The main assumption for this assessment method is that all different types of writing tasks have different success requirements (Davis, 2018; Frey, 2018), so it is necessary to determine the type of task to measure a particular skill and create scoring rubrics that prioritize targeted outcomes, focusing on the aspect that contributes to the task being scored. It is assumed that the rubric used to assess primary characteristics is only appropriate for a single task and cannot be applied to other tasks. This was confirmed by Cohen (1994), who stated that this approach to the rating rubric focuses on one criterion to tailor the rating scale to the task at hand. Compared to holistic and analytical evaluation rubrics, primary trait

scoring is rarely used in studies, so it is the task of this researcher to describe the characteristics of this evaluation rubric following (Frey, 2018). The characteristics are:
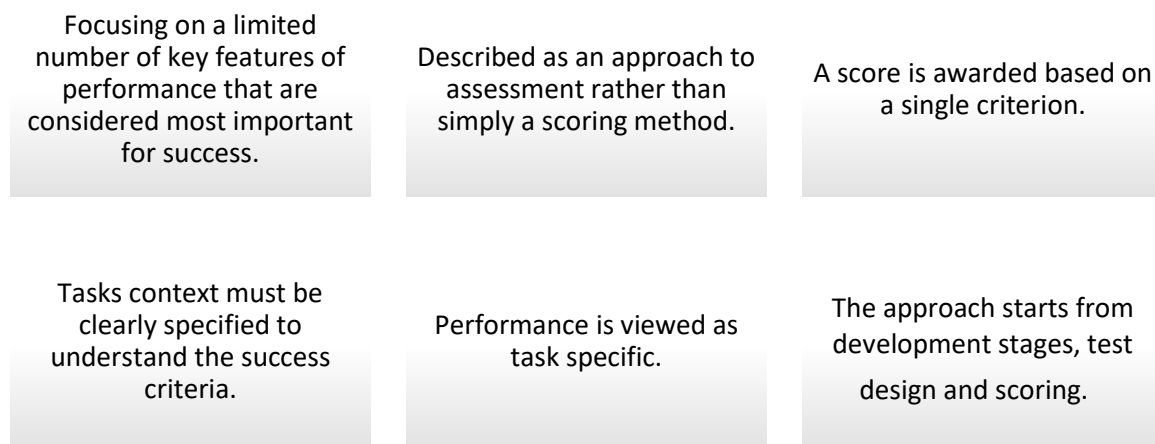
Focusing on a limited number of key features of performance that are considered most important for success.	Described as an approach to assessment rather than simply a scoring method.	A score is awarded based on a single criterion.
Tasks context must be clearly specified to understand the success criteria.	Performance is viewed as task specific.	The approach starts from development stages, test design and scoring.

Figure 2: Characteristics of Primary Trait Scoring Rubric (Frey, 2018)

In addition to characteristics, the focus is on developing a rubric for assessing primary features, based on Davis (2018) Primary Trait Development Model. Selecting the task to be assessed or scored is the first step in developing the primary trait scoring rubric. Here, teachers describe the role required to cover the targeted assessment component. For example, writing an email to a parent is found in the first chapter of the textbook KSSM-CEFR form 5. The item that is thought to be critical to the writing assessment is given priority, and then a prototype of the task is created. This is given to the students, who then create a collection of their writing. Based on the results of the student sample and the theoretical and empirical understanding of the requirements for the effective completion of the task, an assessment criterion is established.

Selecting the task to be tested

⇩

Developing prototype task

⇩

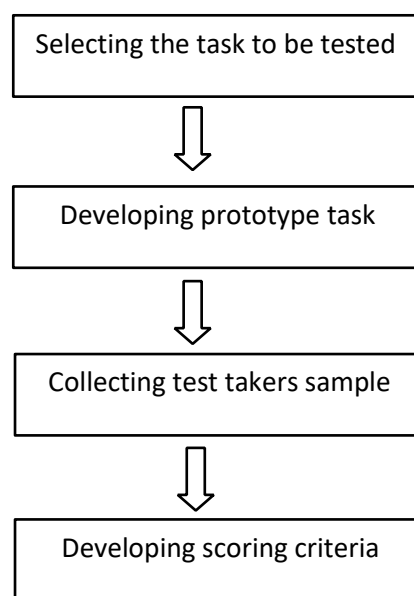Collecting test takers sample

⇩

Developing scoring criteria

Figure 3: Development of Primary Trait Scoring (Davis, 2018)

The strengths and weaknesses of Primary Trait Scoring have been identified by Davis (2018) and can be seen in the table below.

Table 2
Strengths and weaknesses of primary trait scoring rubric

Strengths of Primary Trait Scoring	Weaknesses of Primary Trait Scoring
1. It points out a special feature that is important for a successful performance	1. It requires expertise and a lot of time to create separate scoring materials for multiple tasks.
2. Explicit description of key performance aspects for efficient and targeted guidance for teachers and students.	2. Requires careful analysis of the example of a successful and an unsuccessful performance.
3. Allow teachers to focus on only one aspect	3. The aspect chosen may not represent student performance
4. Focusing on effectiveness in specific contexts gives a more realistic idea of what people can do.	4. It takes time to test the evaluation material with raters.

In the literature where primary trait rubric is used, there are a few studies that have been found. One of the findings shows that the primary trait rubric helps them focus on assessing student essays and is less time-consuming than holistic and analytic assessment in a study by Veloo et al (2018), which examines ESL teachers' preference for the most effective assessment method for testing essay writing in the classroom. It is better to assess students' essays with a clearer and more detailed summary of a student's writing skill for a particular task as well as a specific score (Lloyd-Jones, 1977, cited in Frey, 2018). In a written assignment, the scoring rubric guides raters to focus their attention on a specific discourse element. In the question, scoring is sharpened for the targeted feature and narrowed to the writing task. It is appropriate for classroom assessment and may vary from teacher to teacher (Veloo et al., 2018).

In contrast, one study found that one of the disadvantages of the primary trait scoring rubric is that it is usually difficult for primary raters to focus only on the stated trait because they can often inadvertently include other traits in their rating (Nodoushan, 2014). Primary raters are raters who have experience in scoring essays, and when they analyze students' texts, they are not able to control their expertise. Therefore, it has been proven that most studies use the holistic and analytical scoring rubric only when they use experienced raters in their studies. For inexperienced raters, it may be appropriate to use the primary trait rating rubric, especially for ESL learners. It is helpful to focus on only one factor when reviewing texts. Davis (2018) also reports that the primary trait assessment rubric is well suited for classroom use because teachers can create rubrics based on the trait that they want students to measure.

Primary Trait Scoring is also considered a legitimate topic to be explored because of the need for a very specific rubric for L2 learners that focuses on a specific trait at a specific time in their language learning. To support them in their writing (Weigle, 2010), L2 learners need

a very explicit and comprehensive guide that can provide them with meaningful suggestions for improving their written expression. Second, it is easier for raters to compare their work with that of their peers because they are among students who are also novice and L2 learners. This is because the primary feature assessment provides an explicit definition of the main aspect of success they need to evaluate (Davis, 2018), and this allows them to focus entirely on the desired aspect when evaluating the texts.

Therefore, based on the above explanation, primary trait scoring rubrics can be used by students in Malaysia as they are both novice and L2 learners. The primary trait scoring will allow students to master one aspect at a time as they focus on only one criterion at a time. In terms of SAPA, a new practice in Malaysian classrooms, it is a push for students to assess their progress and they can focus on the criterion they need to assess without mixing it with other criteria. Therefore, primary trait assessment has the potential to be used among student assessors at SAPA because of its characteristics.

Conclusion

In conclusion, the researcher realizes the potential of the primary trait scoring rubric to be used by student raters in SAPA practice compared to other types of scoring rubrics due to its characteristics that suit the needs of L2 learners. The explicit criterion that will be the focus of the primary trait rubric for a particular task will be able to help L2 learners to master a specific aspect that teachers want to focus on. In terms of SAPA, since students are novice raters, a specific scoring rubric that only assesses one aspect at a time will help students to be able to assess their progress effectively thus becoming self-directed learners. This practice will ensure students are capable in the rating process. Apart from that, the implication of this paper is it helps to widen the scope in choosing the suitable scoring rubrics for L2 learners apart from using the commonly used holistic and analytic scoring rubrics as primary trait rubrics might be able to shed light to enhance students learning too. Educators especially teachers in the L2 context can improve their assessment practice with the knowledge of suitable scoring methods to be used in their formative assessment classroom practice. Future research is recommended to develop a primary trait rubric for other learning skills which are speaking, reading, and listening, and identify whether it can be used by students effectively in SAPA.

Corresponding Author
Mohd Effendi @ Ewan Bin Mohd Matore
Senior Lecturer Faculty of Education, Universiti Kebangsaan Malaysia
Email: effendi@ukm.edu.my

References
Shirazi, A. M. (2019). For a Greater Good: Bias Analysis in Writing Assessment. *SAGE Open 9*(1): 215824401882237. http://journals.sagepub.com/doi/10.1177/2158244018822377.

Andrade, H. L., & Du, Y. (2007). Student responses to criteria referenced self-assessment. *Assessment & Evaluation in Higher Education 32* (2): 159–181.

Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education 4*(87). doi: 10.3389/feduc.2019.00087

Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing 29*: 15–24.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability 21*(1): 5–31.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice 25*(6): 551–575. http://doi.org/10.1080/0969594X.2018.1441807.

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review 67*(3): 343–368. http://dx.doi.org/10.1080/00131911.2014.929565.

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy and Practice* 22(4).

Cohen, A. (1994). Assessing language abilities in the classroom. Boston: Heinle & Heinle.

Davis, L. (2018). Analytic, Holistic, and Primary Trait Marking Scales. *The TESOL Encyclopedia of English Language Teaching* 1–6.

Deygers, B., & Van Gorp, K. (2015). Determining the scoring validity of a co-constructed CEFR-based rating scale. *Language Testing 32*(4): 521–541.

Dickinson, P., & Adams, J. (2017). Values in evaluation – The use of rubrics. *Evaluation and Program Planning 65*: 113–116. http://dx.doi.org/10.1016/j.evalprogplan.2017.07.005.

Engelhard, G., & Wind, S.A. (2019). *Invariant Measurement with Raters and Rating Scales: Rasch Models for rater-mediated assessments*. Routledge.

Francis, J. E. (2018). Linking Rubrics and Academic Performance : An Engagement Theory Perspective Linking Rubrics and Academic Performance : An Engagement Theory *15*(1)

Frey, B. B. (2018). Primary Trait Scoring. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc.

Ghalib, T. K., & Al-Hattami, A. A. (2015). Holistic versus analytic evaluation of EFL writing: A case study. *English Language Teaching 8*(7): 225–236.

Hamp-Lyons, L. (1995). Rating Nonnative Writing: The Trouble with Holistic Scoring. *TESOL Quarterly 29*(4): 759.

Holzknecht, F., Huhta, A., & Lamprianou, I. (2018). Comparing the outcomes of two different approaches to CEFR-based rating of students' writing performances across two European countries. *Assessing Writing 37*(April 2017): 57–67. https://doi.org/10.1016/j.asw.2018.03.009.

Idris, M., & Abdul Raof, A. H. (2017). The CEFR Rating Scale Functioning: An Empirical Study on Self- and Peer Assessments. *Sains Humanika 2*(2014): 11–17.

Kaur, P., Zhi, J. (2022). The CEFR-Aligned Curriculum: Perspectives of Malaysian Teachers. *Asian Journal of Research in Education and Social Sciences 4* (1), 138-145. https://doi.org/10.55057/ajress.2022.4.1.13

Kayapinar, U. (2014). Measuring Essay Assessment: Intra-Rater and Inter-Rater Reliability. *Eurasian Journal of Educational Research* (57): 113–135. https://doi.org/10.1016/j.asw.2019.100433.

Martens, K. S. R. (2018). How program evaluators use and learn to use rubrics to make evaluative reasoning explicit. *Evaluation and Program Planning 69*: 25–32.

McNamara, T. F. (1996). *Measuring Second Language Performance*.Pearson Education Limited.

Don, M. Z. (2015). *English Language Education Reform in Malaysia: The Roadmap 2015-2025*. Ministry of Education Malaysia.

Movahedi, N., & Kiasi, G. A. (2021). The Effect of Teacher vs. Learner-Assessment Activities on the Iranian Intermediate EFL Learner's Writing Ability. *International Journal of Research in English Education (2021) 6*(1). http://ijreeonline.com/article-1-467-en.html.

Mumpuni, K. E., Priyayi, D. F., & Widoretno, S. (2022). How do students perform a peer assessment? International Journal of Instruction, 15(3), 751-766. https://doi.org/10.29333/iji.2022.15341a

North, B. (2014). The CEFR in practice. Cambridge: Cambridge University Press.

Ohta, R., Plakans, L.M. & Gebril, A. (2018). Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. *Assessing Writing 38*(August): 21–36. https://doi.org/10.1016/j.asw.2018.08.001.

Othman, N. (2014). The Primary Trait Scoring Method for Classroom-Based Assessment of Students' Direct Writing. *International Journal of Learning and Development 4*(3): 51.

Panadero, E., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education.* 21, 133–148. doi: 10.1080/0969594X.2013.877872

Rahadi, R. A., Tampubolon, M. N., & Hasanah, E. N. (2018). Students Peer-Evaluation Process: A Case Study in Malaysia. *International Journal of Education, Psychology and Counseling 3*(15), 7-16.

Reinholz, D. (2015): The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education 41*(2), 301-315. DOI:10.1080/02602938.2015.1008982

Nodoushan, S. M. A. (2014). Assessing Writing: A Review of the Main Trends. *Studies in English Language and Education 1*(2): 118.

Schunn, C., Godley, A., & DeMartino, S. (2016). The Reliability and Validity of Peer Review of Writing in High School AP English Classes. *Journal of Adolescent and Adult Literacy* 60(1): 13–23.

Sidhu, G. K., Kaur, S., & Chi, L. J. (2018). CEFR-aligned school-based assessment in the Malaysian primary ESL classroom. *Indonesian Journal of Applied Linguistics 8*(2): 452–463. http://ejournal.upi.edu/index.php/IJAL/article/view/13311.

Simons, M., & Colpaert, J. (2015). Judgmental evaluation of the CEFR by stakeholders in language testing. *Revista de Linguistica y Lenguas Aplicadas 10*: 66–77. http://dx.doi.org/10.4995/rlyla.2015.3434.

Trace, J., Meier, V., & Janssen, G. (2016). "I can see that": Developing shared rubric category

interpretations through score negotiation. *Assessing Writing 30*: 32–43. http://dx.doi.org/10.1016/j.asw.2016.08.001.

Veloo, A., Ramli, R., & Khalid, R. (2016). Assessment Practices among English Teachers in Malaysian Secondary Schools. *International Journal for Infonomics 9*(4): 1220–1227.

Veloo, A., Aziz, N. H. A., & Yaacob, A. (2018). The Most Suitable Scoring Method to Assess Essay Writing in ESL Classrooms. *Advances in Language and Literary Studies 9*(4): 19–25.

Wanner, T., & Palmer, E. 2018. Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. *Assessment and Evaluation in Higher Education* 43(7).

Weigle, S. C. (2010). *Scoring procedures for writing assessment*. (J. Charles Alderson & L. F. Bachman, Eds.) *Assessing Writing*. Cambridge: Cambridge University Press. (original work published 2002).

Zou, S., & Zhang, W. (2017). Exploring the adaptability of the CEFR in the construction of a writing ability scale for test for English majors. *Language Testing in Asia 7*(1): 18. http://languagetestingasia.springeropen.com/articles/10.1186/s40468-017-0050-3.