



The Development Of Primary Trait Writing Rubrics For Self-Assessment Activity: A Qualitative Perspective

Nurul Fariena binti Asli¹, Mohd Effendi Ewan bin Mohd Matore^{2*}

¹Faculty of Education, The University of Malaysia (UKM) Bangi, Selangor, Malaysia Email: nurulfariena87@gmail.com

^{2*}Faculty of Education, The University of Malaysia (UKM) Bangi, Selangor, Malaysia Corresponding Email: effendi@ukm.edu.my

Citation: Nurul Fariena Asli & Mohd Effendi Ewan Mohd Matore (2024), The Development Of Primary Trait Writing Rubrics For Self-Assessment Activity: A Qualitative Perspective *Educational Administration: Theory And Practice*, 30(4), 8950-8962

Doi: 10.53555/kuey.v30i4.2934

ARTICLE INFO

ABSTRACT

The alignment of Common European Framework of Reference (CEFR) in the language learning curriculum has put more focus on self-directed learning which is a new practice in the Malaysian classroom. This study delved into developing a type of scoring rubric that is believed to be suitable to be used by L2 learners in self-assessment activities. Since there is no rubric for the intended purpose available to be used in the classroom, this qualitative study was conducted to examine specifically how to construct a suitable scoring rubric (Primary Trait Writing (PTW)) rubric as well as to develop the descriptors in developing the PTW rubric based on the Primary Trait Development Model by Davis (2018). This model was expanded to be more comprehensive as well as to fit the purpose of this study. The study was conducted using samples selected conveniently which involved 15 Form Four students from three local secondary schools in a district in Pahang. Data was collected using a Multi-source Driven Approach consisting of document analysis (curriculum documents and sample essays), expert judgement (expert review using metarubric), and expert intuition. The result was analyzed qualitatively to find the appropriate scale, traits and descriptors that can be used in developing the PTW rubrics for student self-assessment activity purposes. Findings showed that there are four traits (1) content (2) format (3) cohesive device (4) sentence fluency with a simplified version of the descriptors. The results of this study are hoped to assist English as a Second Language (ESL) students in improving their writing skills and providing teachers with an alternative way to conduct effective assessment in the classroom.

Keywords: self-assessment rubric, rubric development, primary trait rubric, metarubric, multi-source driven approach

1. Introduction

In recent years, there has been a growing emphasis within the research community on the utilisation of scoring rubrics as a tool within the criterion-based approach. The quality of each essay is evaluated based on certain criteria specified in the grading rubric, which include coherence, grammatical accuracy, contextual suitability, and other relevant elements. A rubric is a compilation of criteria that encompasses delineations of various degrees of performance (Brookhart & Chen, 2015). It is specifically tailored for an evaluation and delineates the criteria that users are expected to meet. Moreover, it functions as the benchmark against which an assessment will be evaluated (Francis, 2018) and is deliberately utilised to facilitate students in achieving particular objectives. Furthermore, it is a comprehensive compilation of criteria that encompasses thorough elucidations of various levels of performance excellence (Brookhart, 2013). Performance level explanations are not provided alongside the criteria, unlike rating scales. Rating scales often possess a wide-ranging breadth and can be comprehended within the context of a particular setting. The development of a rubric is derived from rating scales, which are further refined to align with the stated objectives of its developers.

Nevertheless, Brookhart (2013) contended that the majority of rubrics were claimed as performance rubrics that fail to accurately depict a performance. A rubric will be determined by two main factors: descriptions of the levels of achievement for these criteria and coherent sets of criteria. This observation indicates that the rubric possesses a descriptive nature rather than an evaluating aspect. The performance must align with the provided description, rather than being solely evaluated and assigned grades. Efficient rubrics should include

suitable criteria and well-crafted descriptions of performance. Consequently, a reliable score or band can be acquired by a rating procedure that utilises a rubric to evaluate performances. Rubrics are utilised to assess the quality of performance, providing feedback and instruction, in contrast to judging methods that do not employ rubrics.

In this study, the evaluation of writing performance was conducted utilising a specially developed writing rubric. Extensive research will be undertaken to ascertain the suitable rubric, encompassing the selection of the rubric type and the delineation of the performance criteria for the rubric. There are two commonly utilised scoring rubrics in research: holistic rubrics and analytic rubrics. According to Weigle (2010), the holistic kind of analysis offers a comprehensive understanding of a written work, while Ghalib and Al-Hattami (2015) argue that the analytic style of rubric concentrates on more specific aspects of the writing. However, an alternative rubric that is relatively underutilised in scholarly literature is the trait-based rubric, which selectively highlights specific writing traits for a given task. To ascertain the appropriate rubric for a study, Weigle (2010) argues that two factors must be considered. Firstly, whether the scale is designed for a specific task or applied to a group of tasks. Secondly, whether the rated script will receive a single or multiple score.

Consequently, the current qualitative study offers fresh insight into how to develop scoring rubrics for self-assessment purposes in the classroom. By adopting the Primary Trait Development Model (PTDM) in developing scale, criteria, and descriptors of the Primary Trait Writing rubrics. Thus, the objective of this study is to develop the scoring criteria, descriptors and performance levels for Primary Trait Writing (PTW) rubrics. Meanwhile, the guiding research questions are (1) How to develop the PTW rubrics which are suitable to be used in Self- Assessment activities for L2 secondary school students in Malaysia? (2) what are the traits and descriptors to develop the PTW rubrics?

2. Literature Review

2.1 Approaches to Rubric Development

The construction of a rating scale can be undertaken through the application of an **intuition-based approach**, which encompasses the integration of expert assessments, committee structures, and experiential approaches. The objective of these methodologies is to establish a measurement tool that effectively captures the competencies and expertise of the software developers (Li & Wang, 2021). In this approach, experts predominantly employ their prior knowledge and comprehension to adapt preexisting descriptors or generate novel descriptions for the scale. Subsequently, the scale is evaluated using language performance samples, and modifications are made to both the scale and criteria until they precisely correspond to actual language performances and satisfy the experts. The aforementioned procedure is conducted gradually to guarantee that the scale exhibits favourable characteristics and capabilities (Davis, 2015; Fulcher, 2012). According to Li and Wang (2021), this particular methodology is commonly utilised in the initial stages of scale development. However, it has been exposed to criticism due to its ability to introduce subjectivity and its lack of empirical validation. Nevertheless, although employing an intuition-based method can serve as a valuable initial step in the construction of a rating scale, it is imperative to supplement it with empirical evidence and a comprehensive development procedure to produce a scale that is both more precise and reliable. The objective of this process is to achieve a balance between the expertise of developers and the need for empirical data and practical relevance. This ensures that the resultant scale is accurate and useful for assessing the intended concept. Teachers commonly evaluate students' language competency inside the classroom setting. Therefore, the incorporation of the intuition-based approach holds significant value in offering guidance for the development of scales (Khamboonruang, 2020).

Furthermore, the utilisation of a **theory-based approach** in the development of rating scales entails commencing with a construct or model that represents the skill under examination and subsequently constructing a scale that aligns with this theoretical framework (Wu, 2021). This methodology is employed across diverse domains, encompassing evaluation and assessment, to ensure that the rating scale is firmly rooted in pertinent theories and offers a methodical framework for comprehending and scrutinising the issue under review. In the field of scale development, there exist four prominent theories that can be employed. These theories include the Four Skills Model, the Models of Communicative Competence, the Theories of Writing, and the Models of decision-making by expert judges (North, 2003; Knoch, 2007). However, relying just on theoretical frameworks would be inadequate for the development of PTW rubrics in this study. Another technique for generating rating scales is the empirically generated approach, which entails creating rating scales using data obtained from observations or experiments, rather than depending on theoretical assumptions or expert judgements. Upshur and Turner (1995) propose using the performance of past test takers to construct criteria for second language competence rating scales that are based on empirical data. This approach ensures that the scales are reliable and valid. The researchers emphasise the lack of explicit instructions in the development of rating scales, leading them to depend on ideas that inadequately capture the intricacies of the language classroom.

In addition, the formulation of a rating scale within a **curriculum-oriented approach** involves the development of a scale that aligns with the specific curriculum and educational objectives. The primary aim of this methodology is to ensure that the rating scale effectively reflects the skills and competencies that are highlighted in the curriculum. The process of developing a rating scale that is aligned with a curriculum-oriented approach involves considering the specific learning outcomes, abilities, and competencies the curriculum specifies. The alignment between the descriptors and criteria of the rating scale and the educational objectives should be ensured. Based on the findings of the study, it is advisable to incorporate relevant concepts into the development of a rating scale to ensure its precision and uniformity. The process involves examining the theoretical underpinnings of the curriculum and utilising them as a framework for developing the grading system. Furthermore, the study emphasises the need to consider the feasibility of the examination while developing a rating scale. Validating the scale is crucial to ensure that it effectively measures the intended skills and competencies. In summary, the development of a rating scale using a curriculum-oriented approach involves aligning the scale with the particular curriculum, considering pertinent theories, and assessing the scale's practicality and validity through validation analysis.

Creating writing rubrics through a **multisource-driven approach** involves considering several perspectives and approaches to create comprehensive and effective rubrics for assessing student work. The integration of various scale development approaches can give rise to a hybrid strategy, commonly referred to as a "multisource" approach. The objective of this strategy is to address the limitations associated with the previous methods (Banerjee et al., 2015). Rubrics are structured evaluation instruments that elucidate writing criteria, augment student performance, and provide precise and consistent assessment. The assessments encompass several key elements, namely the application of criteria to evaluate the product or performance, the utilisation of diverse quality levels, and the implementation of a scoring technique. The establishment of the assignment's aim and desired learning outcomes for students is of utmost importance when developing rubrics. Due to the limitations of the previous approaches, a multisource-driven approach is utilised in this study and it functioned as a conceptual structure for establishing the standards for the rubric.

2.2 Primary Trait Development Model

The primary trait scoring rubric is a rating scale that has proven to be highly successful in measuring writing skills, especially among second language (L2) learners. It was first designed and developed in the early 1970s by Lloyd Jones for the National Assessment of Educational Progress (NAEP), a comprehensive testing scheme for American schools. Following this, researchers have utilised primary trait scoring in many settings, such as the assessment of proficiency in a second language. This scoring technique is based on the underlying principle that each writing activity has its own set of criteria for determining achievement (Davis, 2018; Frey, 2018). Hence, it is crucial to ascertain the nature of the work to assess a particular skill and construct scoring rubrics that give priority to the intended performance, with a specific emphasis on the elements that contribute to the task under evaluation. The scoring rubric for major traits is specific to the work at hand and cannot be generalised to other tasks. According to Boud et al. (1999), the scoring rubric technique was characterised by its emphasis on a singular criterion, which allowed for the customisation of the rating scale to suit the specific work under evaluation.

Davis (2015) offers further elucidation regarding the specific rubric category referred to as the primary trait scoring rubric. In contrast to other rubrics, this particular type does not have the objective of thoroughly defining the performance attributes at every level. Instead, it only concentrates on the particular linguistic events that are crucial for effectively carrying out the task. As described by Frey (2018), there are a few characteristics of the primary trait scoring rubric. Firstly, it places a limited emphasis on a specific set of performance attributes that are considered vital for attaining success. Secondly, it assigns a score based on a singular criterion. Lastly, it is characterised as an assessment technique rather than a simple scoring system. Performance is commonly understood to be task-specific. The fundamental approach to trait scoring comprises the many stages of development, the design of the test, and the process of scoring. Furthermore, Davis (2018) has delineated the sequential phases included in the development of a primary trait rubric, as illustrated in Figure 1.

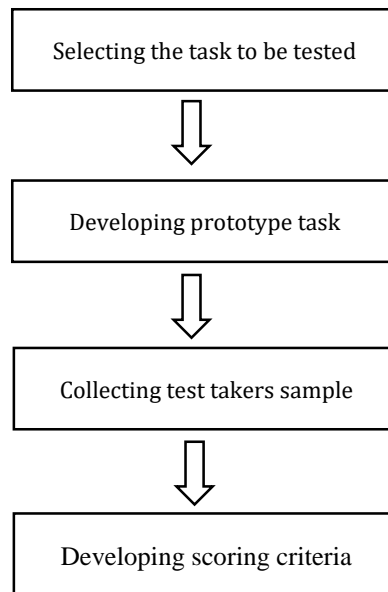


Figure 1: The Development of Primary Trait Rubric

To develop a primary trait scoring rubric according to Davis (2018), the first step is to select the particular task that will be assessed or examined. During this phase, educators will ascertain the precise task necessary to address the particular aspect under assessment. When assessing writing, the crucial aspect is the emphasis, which leads to the creation of a prototype of the work. The allocation of resources will be conducted among the students, and a compilation of their written compositions will be compiled. A set of grading criteria will be developed by considering the observed performances of the students' sample, as well as the theoretical and empirical knowledge of the requirements for completing the assignment.

2.2 Self-Assessment in the Curriculum

The study will utilise the self-assessment commonly referred to as SA, as a means for students to do writing tasks. According to Zuraidah (2015), Malaysia is presently in the second stage of the English Language Roadmap, which necessitates the adoption of assessment protocols that foster student autonomy and self-directed learning. Hence, it is assumed that students possess either limited knowledge of SA or have not yet been acquainted with the activity at that particular phase. The assertion presented is based on contemporary scholarly investigations on School-Based Assessment (SBA), which suggest a dearth of Systematic Assessment for Progress in Achievement (SAPA) that has been implemented within educational establishments (Idris & Abdul Raof, 2017; Sidhu et al., 2018). This section will examine SA (Self-Assessment) to enhance understanding of each activity.

The new English language learning curriculum in Malaysia emphasises that for students to meet predetermined goals and become successful independent learners, they must develop the ability to monitor and control their thoughts, feelings, and behaviours (Panadero & Romero, 2014). According to Zuraidah (2015), educators should incorporate self-assessment (SA) as a formative assessment in the classroom due to its potential to enhance students' self-regulation in the learning process. Self-assessment (SA) is a method of evaluating one's work and learning by examining the clearly defined goals or criteria for success. Subsequently, they assess the merits and limitations and implement modifications to enhance their job (Andrade & Du, 2007). As to Taras (2010), self-assessment (SA) can be defined as the comprehensive evaluations given by learners regarding their work.

Andrade and Du (2007) proposed a pedagogical three-step process in their study. The initial stage entails the instructor delivering the intended performance standards to the students. During the subsequent phase, students proceed to do their designated task and assess their work by employing the allocated rubric. Ultimately, during the third phase, students engage in the process of revising and improving their work by incorporating the feedback they have obtained through self-assessment. The steps outlined in this study closely resemble those suggested by Brown et al. (2015). The self-assessment process consists of three essential principles and steps: (1) Establishing expectations, (2) assessing oneself, and (3) implementing improvements. Hence, a crucial step in conducting self-assessment (SA) involves the teacher setting explicit objectives for a particular activity and providing students with a comprehensive explanation of each criterion that will be utilised in the assessment procedure.

To guarantee that students possess a comprehensive understanding of the precise criteria they are required to achieve, it is imperative to engage in an exhaustive conversation with them discussing the parameters of a rubric in a writing evaluation work. Afterwards, in the second stage, students will write their essays according to the given task and assess themselves using the predetermined criteria. If individuals do not achieve their desired goals, they might acknowledge the mistakes they have made. In the last phase, students formulate the final assignment and assess it in relation to their instructor.

2.3 6+1 Trait Writing Model

The learning technique known as the 6+1 Trait Writing Model involves the comprehensive teaching, modelling, and evaluation of the writing process and it was developed by Spandel and Stiggins (1990). The emphasis lies on seven fundamental attributes of excellent writing: concepts, structure, tone, vocabulary selection, coherence of sentences, conventions, and delivery. The methodology provides a compilation of rubrics that outline the standards for assessing the quality of student writing and tailoring feedback for students. These rubrics have been subjected to thorough testing in the field, are based on research, and have been specifically designed to simplify their use in various writing genres, such as informative/explanatory, argumentative, and narrative writing. The objective of this notion is to support educators in providing effective feedback to students and promoting consistency across all grade levels. Moreover, it empowers students to assess their writing by considering the unique characteristics establishing a common language and establishing criteria for writing guidance. This model, known as the 6+1 Trait Writing Model, is widely utilised in the realm of education to improve students' writing skills and provide specific feedback to facilitate their development as writers.

2. Methodology

3.1 Research Design and Participants

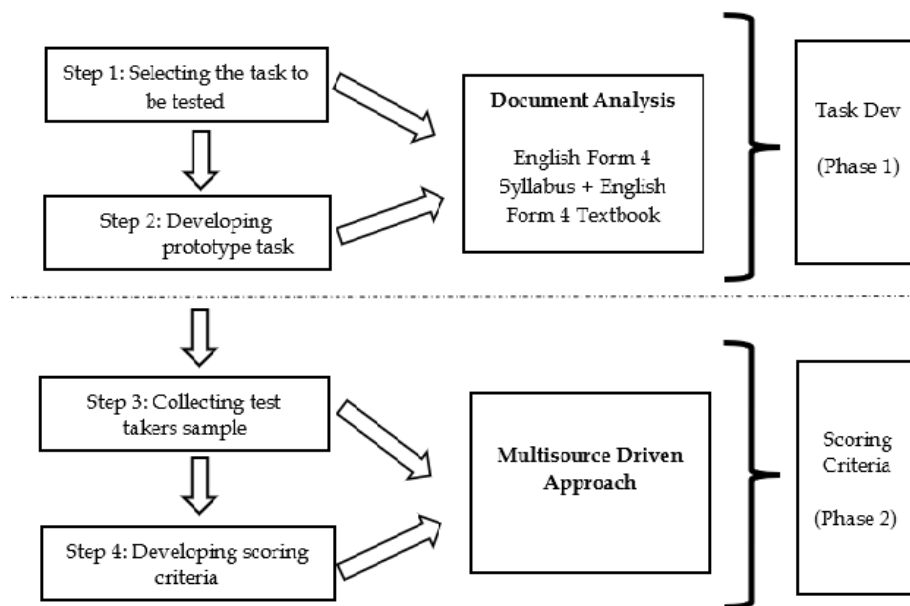


Figure 2 Expanded Primary Trait Development Model

This stage aimed to design and construct the draft of the PTW rubrics qualitatively by analyzing various qualitative data resources. In this study, the Primary Trait Development Model by Davis (2018) was expanded, and it was shown in the figure above. The PTDM was divided into two phases and for each phase, the data collection method was specified. The two phases in this stage are (a) phase 1- task development phase (step 1 and step 2), (b) phase 2- scoring criteria development phase (step 3 and step 4).

3.1.1 Phase 1 (Task Design)

Two steps were taken to ensure that the writing tasks for the self-assessment activity aligned with the curriculum of the English subject and its target population. The researcher conducted a document analysis on the required Form 4 English syllabus, the English Textbook for Form 4, and the Guidelines for Examiner manual from the Malaysia Examination Syndicate as part of the initial investigation. The objective of this study was to establish the validity of the tasks through a comparative analysis of the curricular materials pertaining to the subject matter. When creating a task to evaluate language proficiency using CEFR as a benchmark, the task specifications can be derived from appropriate descriptors for each level (Harsch & Sayferth, 2020). However, for this particular study, the task specifications were specifically designed for the A2 level. The selection of the A2 level challenge was based on its ability to accommodate students with varying degrees of

skill to address the task. Table 1 presents the test specification table for the writing activities that have been specifically prepared for this assessment.

Table 1 Writing Task Specification Table

Part/number of questions	Task Type	Textbook Unit and Theme	Number of words	Intended CEFR level	Objective/aim
Part 1 / 1	Short Communicative Message (e-mail)	Unit 2: Ready for Anything Theme: People and culture	80 words	A2	-This task requires a personal and concrete response on general social and everyday matters (Examination Syndicate, 2020)

The purpose of the second step was to verify the writing task by incorporating the feedback provided by experts through the expert validation form. Two individuals who possessed expertise in teaching Writing Assessment courses at two public universities and were actively engaged in the development of teaching materials for English as a Second Language (ESL) courses were chosen. The validation process is of utmost importance to establish consensus, which refers to the agreement reached via the assent of experts (Morales et al., 2018), and to guarantee the task's content validity. The questions in the validation form encompassed the clarity of the language, the suitability of the assignment for the intended audience, and the appropriateness of the tasks for the chosen subjects. Regarding the aspect of face validity, the task was developed based on the SPM (high-stake exam) essay paper, which the students were already acquainted with. Finally, the feedback provided by the experts was utilised to adjust the task. After the completion of the prototyped task, the subsequent stages 1 and 2 in the PTDM were effectively executed.

3.1.1 Phase 2 (Scoring Criteria Development)

Subsequently, the development phase of the scoring criteria, which corresponded to phases 3 and 4 in PTDM, was then the main focus. The sample comprised 15 form four students from three secondary schools in Pahang, who were tasked with writing essays based on a predetermined task. The researcher employed a purposive sampling methodology to pick participants for this study since all participants were chosen based on their ability to provide comprehensive and essential information for the development of the PTW rubrics. The researcher employed purposive sampling to deliberately choose individuals who possessed the specific experiences, viewpoints, and characteristics necessary for collecting data on the primary concept investigated in the study (Creswell & Plano Clark, 2018).

The selection of these students was based on their teacher's assessment of their language proficiency, which was categorised into three levels: (a) high proficiency, (b) intermediate proficiency, and (c) low proficiency. The researcher retrieved and combined data from various proficiency levels to guarantee that the characteristics, standards, and ratings for PTW rubrics were appropriate for usage by every student. The participants consisted of students selected based on the researcher's decade-long experience as an English teacher, which yielded a substantial amount of data for this study. After gathering the example essays, the study employed a Multisource Driven Approach. In this study, the descriptors for each characteristic were extracted from students' written writings. These sources successfully addressed the limitations of each method and offered the advantage of triangulation.

3.2 Instrumentation

This stage involved the utilisation of two instruments, namely the writing task and a metarubric (Steven & Levi, 2005). The metarubric has been established by experts for the aim of validation. Its purpose is to assess the clarity and meaningfulness of the descriptions, as well as to determine if they accurately represent varying levels of competence. Additionally, the rubrics are designed to be interpreted meaningfully by the intended participants. Experts wrote their comments on the designated form together with the metarubric.

Table 2 Metarubric

Rubric Part	Evaluation Criteria	Absolute Agreement Percentage
1)The trait	1a) Does the trait cover important part of students' performance?	
	1b) Is the trait clear?	
	1c) Is the trait being taught on your teaching for essay writing?	
2) The description of levels of performance	2a) Do the descriptions match the trait?	
	2b) Are the descriptions clear and different from each other?	
	2c) Is there a clear basis for assigning points for the trait?	
	2d) Since it is a five-level rubric, are the descriptions appropriately and equally weighted across the levels?	
3) The scale level	3a) Do the descriptors under each level truly represent that level of performance?	
	3b) Are the scale labels (0-4) encouraging and quite informative?	
	3c) Does the rubric have a reasonable number of levels for the stage of the student and the complexity of the task?	
4) The overall rubric	4a) Does the rubric clearly connect to the learning outcomes that it is designed to measure?	
	4b) Can the rubric be understood by external audiences (avoid jargon and technical language)	
	4c) Does it reflect teachable skills?	
	4d) Is the rubric appropriate for the conditions under which the task was completed?	
	4e) Does the rubric emphasize the appraisal of individual performance?	
5) Fairness and sensibility	5a) Does it look like the rubric will be fair to all students and free of bias?	
	5b) Does it look like it will be useful to students as performance feedback?	
	5c) Is the rubric practical given to the kind of task?	
	5d) Does the rubric make sense to the reader?	

3.3 Data Collection

The collection of students' sample essays was conducted in accordance with the writing task. The most important traits appeared and were chosen as the primary components which need to be focused on when creating PTW rubrics for a certain writing task. The guideline for this stage was based on the Primary Trait Development Model (PTDM) proposed by Davis (2018) and the Multisource Driven Approach. After the development of the rubric, a metarubric was employed to assess the rubric by four experts. The overall agreement percentage among the experts was then utilised to establish consensus and validation about the traits and descriptors.

Table 3 Breakdown of Multi-source Driven Approach

Approaches	Sources
Intuition Based	Teacher's experience
Theory Based	6+1 Trait Model including 6+1 self-assessment rubrics
Empirically Derived	Performance data-driven: targeted audience sample scripts
Curriculum Oriented	Syllabus, Textbook

3.4 Expert Review Judgement

The expert review judgement was conducted at this step using a metarubric derived from Stevens and Levi (2005). According to Stevens and Levi (2005) and Arter and McTighe (2001), a metarubric refers to a rubric that is employed to assess the effectiveness of a proposed rubric. Therefore, it was implemented to engage experts in order to obtain their consensus and validation about the constructs of the established rubric. Through the utilisation of the metarubric, specialists would alternate between the rubric and the metarubric criteria, so enhancing and refining certain aspects of the rubric. The utilisation of the yes/no element in all

metarubrics facilitates a rapid assessment of the fundamental components involved in rubric building. The evaluation of the overall quality of a rubric involves the utilisation of five distinct components: a) the dimensions, b) the descriptions, c) the scale, d) the overall rubric, and e) fairness and sensibility. Furthermore, there are evaluation criteria for each component of the rubrics. The inter-rater agreement was determined using an absolute agreement, as metarubric relies on yes and no responses from experts. According to the Standards for Educational and Psychological Testing (AERA/APA/NCME, 1999), the criterion for assessing the adequacy of inter-rater agreement in deciding the appropriateness of ratings for consequential use is as follows:

Table 4 Rule of thumb for determining inter-rater agreement

Agreement Summary Statistic	High	Minimum	Comment
Absolute Agreement	90%	75%	There should be no ratings more than 1 level apart. If there are more than 5-7 rating levels, an absolute agreement level closer to 75% would be acceptable, but exact and adjacent agreement should be close to 90%

The results in Table 4 show the absolute agreement percentage that was calculated using the input presented by all four experts. Based on the facts supplied in the table, it can be deduced that all four experts expressed positive replies to a certain topic, as indicated by a value of 100%. If the proportion is 75%, it signifies that just three individuals provided an affirmative response, while a percentage of 50% implies that two individuals provided an affirmative response. Ultimately, if the calculation produces a result of 25%, it signifies that only one expert provided a positive response. Furthermore, the experts offered some suggestions for improving the rubrics, one of which was the suggestion to integrate an additional rubric focused on sentence fluency.

4 Results and Discussion

4.1 Findings on the Scoring Criteria, Descriptors And Performance Levels For PTW Rubrics

This part addresses the initial purpose of the study by addressing two research questions (1) how to develop the PTW rubrics which are suitable to be used in Self- Assessment activities for L2 secondary school students in Malaysia? and (2) what are the traits and descriptors to develop the PTW rubrics? In this study, the researcher employed the Multi-source Driven Approach to develop rubrics for PTW. The researcher gathered data from various sources, including language performances through students' writing samples, curriculum documents such as the English Language Form 4 syllabus and textbook, the 6+1 trait model and its existing rubrics examples from the literature, and expert intuition.

The analysis of data involved utilising the literature study, rubric samples, and curriculum resources to get information on the features. Additionally, the criteria were determined by incorporating expert intuition from teachers and writing samples from students. To ensure the accurate interpretation and application of the rubrics before progressing to the next stage, a select group of language and writing assessment experts are requested to evaluate the contents of the draft scale using metarubric. Upon the conclusion of this phase, the first research question was addressed.

Regarding the second research question, the document analysis revealed that the four criteria for assessing writing, as proposed by CEFR, are Content, Communicative Achievement, Organisation, and Language. These criteria are based on the CEFR assessment grid. According to Harsch and Seyferth (2020), the four criteria mentioned have undergone a comprehensive validation process by a test development team for the Common European Framework of Reference for Languages (CEFR), thereby establishing their indisputable applicability. This study employed the four criteria as a framework for selecting the traits for PTW rubrics. Given the broad nature of the four criteria and the limited scope of the major characteristic rubric, an analysis was undertaken to ascertain the appropriate trait for each criterion.

In addition, 6+1 trait rubric samples were utilised for self-assessment in order to identify the specific feature of the targeted PTW rubric. In addition, the primary material consisted of sample essays submitted by students. The sample essays written by the students were sorted into three competency levels using Brookhart's (2013) approach to developing a classroom rubric. This allowed for the observation of variations in the sample essays. The findings from the analysis are presented in Table 1.

Table 5 Findings from students' sample essays

Traits/ Level	TP 1 and TP 2	TP 3 and TP 4	TP 5 and TP 6
Content	Do not answer the questions or questions are answered but a) do not have supporting details b) only have 1 supporting detail	Answer the questions but have 1 and 2 supporting details only	Answer the questions and have 3 or more supporting details
Communicative Achievement = Format	Do not use any elements of the format Only have 1 element (body of the email)	Use 1 or 2 elements of the format (without salutation or signing off)	Use 3 and 4 elements of the format
Organization = Cohesive Device	Do not use any cohesive device Use only simple connectors (and, but, so)	Use 1 or 2 common cohesive Device (Firstly, Lastly)	Use more than 2 cohesive devices (moreover, nevertheless)
Language = Sentence Fluency	Write below than 3 sentences without repeating the same word at the beginning.	Write between 3-6 sentences without repeating the same word at the beginning.	Write more than 8 sentences without repeating the same word at the beginning.

The table presents the results on the strengths and shortcomings of students' writing within each category. Therefore, it was determined that the traits for PTW rubrics, which may be used for self-assessment, would be content, format, coherent device, and sentence fluency. In addition, the data shown in **Table 2** was utilised to generate descriptors for each characteristic. It was compiled and shown in the table below.

Table 6 Description of the traits

Traits	Measures	Descriptions
Content	Number of responses	Total of responses based on the questions
Format	Number of correct formats used	Total number of the format for email being used 1-Salutation 2-Greetings 3-Body of email 4-Signing Off
Cohesive Devices	Number of cohesive devices	Total number of cohesive devices used
Sentence Fluency	Number of sentences used	Total number of sentences being used that start differently

After the development of the traits and descriptors, the performance levels of the rubric were established. The developed rubrics are intended for students of all skill levels, including those with poor ability, and they share the same focus as a study by Uludag & McDonough (2022). The purpose of this study was to validate a rubric designed for EAP students. Four categories—content, organisation, source utilisation, and language use—with a score range of zero to four were included in the pilot-tested and amended rubric. A zero was employed to accommodate students with limited ability, as some of them did not respond to the questions at all. This phenomenon was also observed in the sample of essays submitted by students in the context of this study.

4.2 Discussions on the Selected Scoring Criteria, Descriptors and Performance Levels for PTW Rubrics

The development of PTW rubrics commenced with the development of an assessment task that was employed in conjunction with the established rubrics. The assessment of the task was customised to align with the specifications specified in part one of the SPM exam question. The assignment is to reply to an email that

requests a response, so this is a written interaction question. The objective of the Self-Assessment Grid (SAG) from CEFR is I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something (Council of Europe, 2020).

After undergoing validation by experienced teachers and lecturers, the assessment task was then administered to the students. The collected samples were categorised into three levels of competency to obtain a comprehensive understanding of the task's output, as suggested by Arter and McTighe (2001), Brookhart (2013), and Nurfirdawati (2016). The emphasis was placed on the primary components of CEFR, namely content, communicative achievement, organisation, and language, in order to enhance attention and structure. Given that the rubrics were PTW, the researcher had to provide more precise details. The most prominent aspect that requires attention in students' essays is the employment of coherent devices in the content style and the fluidity of sentences. Following the application of Davis's (2018) primary trait formation model, all four parts were subsequently designated as traits. After identifying the features, a rating scale and criteria were created for each treatment utilising a multi-driven approach.

In addition, this study has expanded upon the Primary Trait Development Model proposed by Davis (2018), so augmenting its comprehensiveness and laying the groundwork for future investigations in this field. The model exhibits a high degree of simplicity and necessitates further expansion, as depicted in Figure 2. In the current study, the PTDM was partitioned into two distinct stages. The initial phase, referred to as the Task Development Phase, encompassed two key steps: step 1, which involved the selection of the task to be tested, and step 2, which entailed the development of a prototype task. Concurrently, the subsequent stage was designated as the Scoring Criteria Development Phase, encompassing PTDM's third step (the collection of a sample of test takers) and the fourth step (the development of scoring criteria). The initial stage involved the execution of document analysis, whereas the subsequent stage employed the multisource-driven approach. This study has made a valuable contribution to future research by offering explicit guidelines for the development of a scoring rubric for primary trait types, through the expansion of the PTDM.

It is crucial to bear in mind that the rubrics are specifically created for assessment and should be easily understandable by the intended recipients. It is imperative to ensure that the vocabulary statement is both comprehensible and intelligible to the students. The Malaysian classroom has implemented the CEFR, and scholars have acknowledged the need to include self-assessment activities as a primary goal. The objective of this method is to foster the development of learners who possess the capacity to assume accountability for their learning and possess the ability to recognise their strengths and limitations. At this juncture, the researcher decided to integrate the involvement of students and proceeded to undertake a preliminary study, sometimes referred to as pre-trialing, before commencing the pilot study. In addition, the pre-trialing phase was undertaken to evaluate the face validity of the PTW rubrics. A cohort comprising 20 students exhibiting diverse levels of proficiency was chosen to partake in the initial trial. Following the comments supplied by the students, subsequent improvements were implemented.

If there is still ambiguity in the qualities' description (Brookhart, 2013), student participation in the rubric construction process is essential, and it is necessary to hear the students' opinions of the rubrics (Deygers & Van Gorp, 2015). Hence, within the framework of this research, the incorporation of a co-constructed rubric was also employed to authenticate the rubrics. The selection of rating scale categories was determined by analysing the replies obtained from students' sample essays. Given that these rubrics must also accommodate students with lower proficiency, the category was initially set to 0, as the sample essays indicate that low-proficiency students receive a score of 0 for a certain feature. Following the research conducted by Uludag and McDonough (2022), the use of zero in their study is also justified as a means to accommodate students with low proficiencies. The qualities selected for the rubrics were the subject of more discussion.

All assessments had to be in accordance with the guidelines from the CEFR since it served as the framework for language learning. The Common European Framework of Reference for Languages (CEFR) encompasses grading systems that offer curriculum planners, lecturers, and teachers a comprehensive framework as reference points. Nonetheless, the CEFR scales present difficulties for language assessors in their use due to the absence of uniformity, accuracy, and relevance of the current descriptors (Simons & Colpaert, 2015; Zou & Zhang, 2017). However, according to North (2014), the scales possess a significant level of generality, enabling them to be applied in diverse contexts without constraining the user's capacity to modify and adjust student learning. Furthermore, it should be noted that the CEFR scales are specifically intended to assess competencies and serve a distinct function compared to assessment scales (Holzknecht et al., 2018). It is important to note that the selected traits for assessing writing in this study were based on the CEFR criteria (Task Fulfilment, Communicative Competence, Organisation and Language) established by the Council of Europe in 2001. However, these traits were specifically tailored to the assessment task and rubrics used in this study. These findings are consistent with the research conducted by Harsch and Seyferth (2020), who claimed that the four

criteria mentioned have been thoroughly validated by a team responsible for developing tests for the Common European Framework of Reference for Languages (CEFR), thus confirming their undeniable appropriateness. The main factor to assess in a writing exercise was the substance and quality of the student's responses, specifically concerning the **Content** trait. The objective of this is to ensure that students possess the ability to furnish responses to the inquiries posed in the assessment. The students' comprehension of the writing task will be showcased through their diverse responses, enabling the evaluation of their proficiency in particular skills. In the context of self-assessment, students have the opportunity to develop an understanding of their limitations and talents, thus enhancing their performance in subsequent pursuits. Meanwhile, the selection of the **Format** trait was based on its role as a guiding concept for determining the appropriate structure of the responses. The CEFR outlines a component of communication skills for students. For example, when the task involves composing an email or an informal letter, students are expected to provide responses that align with the particular genre of writing. In the case of a formal letter, it is imperative to adhere to the prescribed structure and norms that are typically associated with such written communication. The objective is to educate prospective employees to deliver appropriate and precise responses that adhere to the correct language style (register), tone, and specific task demands. To enhance the process of students' self-assessment, a format was chosen that emphasises user-friendliness and promotes self-assessment, rather than emphasising register and tone. The use of a primary trait rubric type serves to improve the level of specificity.

Apart from the **Content** and **Format** characteristics, the CEFR also highlights the significance of using linkers to enhance essay writing's arrangement. Hence, the **Cohesive Device** was chosen as the trait in the rubrics to enhance students' proficiency in employing linkers beyond basic ones such as "and," "but," and "so." The utilisation of cohesive devices, such as "next," "after that," and similar terms, is also employed to cater to students with average to higher proficiency levels in their works. Regarding **Sentence Fluency**, it was selected due to the prevailing consensus among experts that students, particularly those with low proficiency, often struggle to identify their grammatical problems. Consequently, **Sentence Fluency** emerges as a very acceptable trait for students to engage in self-assessment. Additionally, this is based on the 6+1 attribute writing rubrics that were previously employed for self-assessment.

In summary, this study has extended the existing Primary Trait Development Model and through thorough investigation, the traits, categories, and descriptors of the PTW rubrics have been selected. Hence, this phase provided guidance for writing assessment researchers who opt to construct the Primary Trait Writing rubric for their studies, as well as for educators who wish to create their self-assessment instruments for their classroom activities.

6 Conclusion

In conclusion, this qualitative study successfully developed four Primary Trait Writing rubrics based on four traits (Content, Format, Cohesive Device and Sentence Fluency). The Primary Trait Development Model was used as a guideline, and it was expanded to suit the demand for this study. The multi-source-driven approach used had shown triangulation in the development of the PTW rubrics as various approaches were utilised. It can be seen that the sample essays from targeted participants yielded the main findings primarily the descriptors of the rubrics. In addition to that, expert judgements were gauged to determine the validity of the content of these rubrics.

7 Research Implications, Limitations, and Recommendations

The outcome of this study has important educational implications for verifying scoring rubrics used for self-assessment to evaluate and enhance students' writing proficiency. In assessing students' writing skills, the need to choose a suitable type of scoring rubric as well as developing it accordingly to targeted participants is undeniably crucial. After the rubric has been developed, there is a need to ensure that it achieves its construct validity so that it measures what it has to measure based on the experts' judgements as well as the students. In addition to that, theoretical implications can be seen with the expansion of the previous Primary Trait Development Model. The added elements to the previous model are needed to ensure the comprehensiveness of the development of the rubrics. This study also has a few limitations as the primary trait scoring rubric only focuses on specific types of assessment tasks (email) thus, it is recommended that this type of rubric be developed for another type of assessment (informal letter). In addition to that, the study also lacks empirical data. Thus, a combination of qualitative and quantitative methods is recommended to yield better results in developing and validating the developed rubrics.

8 References

1. AERA, APA and NCME. (1999). *The Standards for Educational and Psychological Testing*. Aera Publication, Washington DC.
2. Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment Evaluation High Education*, 32, 159–181. doi: 10.1080/02602930600801928
3. Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin Press.
4. Banerjee, J., Yan, X., Chapman, M., & Elliott, H. (2015). Keeping up with the times: Revising and refreshing a rating scale. *Assessing Writing*, 26, 5–19. <https://doi.org/10.1016/j.asw.2015.07.001>
5. Boud, D. & Cohen, R. & Sampson, J. (1999). *Peer Learning and Assessment*. *Assessment & Evaluation in Higher Education*, 24, 413–426. 10.1080/0260293990240405.
6. Brookhart, S.M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy and Practice* 20(1): 69–90.
7. Brookhart, S. M. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics, *Educational Review*, 67:3, 343–368, DOI: 10.1080/00131911.2014.929565
8. Brown, G., Andrade, H., and Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assess. Educ.* 22, 444–457. doi: 10.1080/0969594X.2014.996523
9. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*, Council of Europe Publishing, Strasbourg. www.coe.int/lang-cefr.
10. Creswell, J. W. & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.
11. Davis, L. (2015). Designing and using rubrics. In J. D. Brown & C. Coombe (Eds.), *The Cambridge guide to research in language teaching and learning* (pp. 238–246). Cambridge University Press.
12. Davis, L. (2018). Analytic, holistic, and primary trait marking scales. *The TESOL Encyclopedia of English Language Teaching*, pp. 1–6. New York, United States: John Wiley & Sons, Inc.
13. Deygers, B. & Van Gorp, K. (2015). Determining the scoring validity of a co-constructed CEFR-based rating scale. *Language Testing* 32(4): 521–541.
14. Francis, J. E. (2018). Linking rubrics and academic performance: An engagement theory perspective. *Journal of University Teaching & Learning Practice*, 15(1). <http://ro.uow.edu.au/jutlp/vol15/iss1/3>
15. Frey, B. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. 10.4135/9781506326139.
16. Fulcher, G. (2012). Scoring performance tests. In G. Fulcher & F. Davison (Eds.), *The Routledge handbook of language testing* (pp. 378–392). Routledge.
17. 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.
18. Holzknicht, 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, 57–67. <https://doi.org/10.1016/j.asw.2018.03.009>
19. Li, J. & Wang, Q. (2021). Development and validation of a rating scale for summarization as an integrated task. *Asian-Pacific Journal of Second and Foreign Language Education* 6(1).
20. North, B. (2014). *The CEFR in practice*. Cambridge: Cambridge University Press.
21. North, B. (2003). *Scales for rating language performance: Descriptive models, formulation styles, and presentation formats*. TOEFL Monograph 24. Princeton: Educational Testing Service.
22. Nurfirdawati, M.H. (2016). *Pembangunan dan penentusahan rubrik pentaksiran prestasi bagi mentaksir projek reka bentuk seni bina di politeknik Malaysia* [Unpublished doctoral dissertation]. Universiti Perguruan Sultan Idris.
23. Harsch, C., & Seyferth, S. (2020). Marrying achievement with proficiency – Developing and validating a local CEFR-based writing checklist. *Assessing Writing*, 43, 100433. <https://doi.org/10.1016/j.asw.2019.100433>
24. 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.
25. Khamboonruang, A. (2020). *Development and Validation of a Diagnostic Rating Scale for Formative Assessment in a Thai EFL University Writing Classroom: A Mixed Methods Study* [Unpublished doctoral dissertation]. The University of Melbourne.
26. Knoch, U. (2007). *Diagnostic writing assessment: The development and validation of a rating scale* [Unpublished doctoral dissertation]. University of Auckland.
27. Morales, J., Montes, R., Zermeño, N., Duran, J., & Herrera, F. (2018). The use of fuzzy linguistic information and fuzzy Delphi method to validate by consensus a questionnaire in a Blended-Learning environment. In *Communications in computer and information science* (pp. 137–149). https://doi.org/10.1007/978-3-319-91479-4_12

28. Panadero, E. & Romero, M. (2014) To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy, *Assessment in Education: Principles, Policy & Practice*, 21:2, 133-148, DOI: 10.1080/0969594X.2013.877872
29. 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.
30. Simons, M., & Colpaert, J. (2015). Judgmental evaluation of the CEFR by stakeholders in language testing. *Revista de Linguística y Lenguas Aplicadas* 10: 66–77. <http://dx.doi.org/10.4995/rlyla.2015.3434>.
31. Spandel, V. Stiggins, R. (1990). *Creating writers. Linking assessment and writing instructions*. New York: Longman.
32. Stevens, D. D., & Levi, A. J. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning* (2nd ed.). Sterling, VA: Stylus Publishing.
33. Taras, M. (2010). Assessment for learning: assessing the theory and evidence. *Procedia - Social and Behavioral Sciences*, 2(2), 3015–3022. <https://doi.org/10.1016/j.sbspro.2010.03.457>
34. Uludag, P. & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing* 52.
35. Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. <https://doi.org/10.1093/elt/49.1.3>
36. Weigle, S.C. (2010). *Scoring Procedures for Writing Assessment*. J. Charles Alderson & L. F. Bachman (Eds.) *Assessing Writing*. Cambridge: Cambridge University Press.
37. Wu, X. (2021). Establishing an operational model of rating scale construction for English writing assessment. *English Language Teaching*, 15(1), 16. <https://doi.org/10.5539/elt.v15n1p16>
38. 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>.
39. Zuraidah, M.D. (2015). *English Language Education Reform in Malaysia: The Roadmap 2015-2025*. Ministry of Education Malaysia.