The Appropriacy and Applicability of English Assessment Against CEFR Global Scale: Teachers' Judgement

NURUL FAREHAH MOHAMAD URI
Universiti Kuala Lumpur British Malaysian Institute (UniKL BMI)
nfarehah@unikl.edu.my
&
Universiti Kebangsaan Malaysia

MOHD SALEHHUDDIN ABD AZIZ
Universiti Kebangsaan Malaysia

ABSTRACT

This study intends to identify the appropriateness of the writing and reading assessment items of the current English syllabus according to the targeted CEFR levels set by the Ministry of Education and to find out English teachers' familiarity with CEFR writing and reading scales by suggesting suitable CEFR levels for writing and reading assessment items. An assessment checklist was used to collect the data involving 331 secondary school English teachers. Results show that most of the assessment items are found to be suitable and applicable to be used in the CEFR. This study also proves that the teachers were aware of the six levels of CEFR global scale because they were able to identify the suitable CEFR level that matches against CEFR target level set by the Ministry of Education. In conclusion, the Ministry of Education, Malaysia should consider to continue using some part of the current English syllabus and assessment instead of recommending its total abolishment.

Keywords: CEFR in Malaysia; Teachers' judgement; CEFR alignment; School based assessment; applicability

INTRODUCTION

The adoption of CEFR onto education system needs to be comprehensive and involves a number of major changes. According to Byram, Gribkova and Starkey (2002), the adoption of CEFR should involve changes to several components of the education system which include the syllabus, teaching approach, teaching materials, learning outcomes as well as assessment and evaluation. The adoption and implementation of CEFR is incomplete and in fact, invalid without aligning the framework against English syllabus and assessment. Henceforth, calibrating the syllabus with CEFR shall be the first main step in the implementation process. This is important as it checks and ensures the syllabus is carefully mapped and aligned with the six levels of CEFR descriptors before it can be claimed as 'CEFR aligned English syllabus.

CEFR alignment against assessment is another important process of CEFR implementation. CEFR six-level global scale is purposely designed to be open, dynamic and flexible to help language users mapping out the CEFR to their assessment and syllabus. Any language user or teacher could easily map his or her test or assessment against CEFR because the Council of Europe has published a manual for aligning language tests to CEFR in 2003. As a result, there are abundant of international and local assessments which have been aligned to CEFR. Among the first to align their language tests against CEFR are major international testing agencies (Little 2011); IELTS and TOFEL are amongst the well-established and popular international language tests which are now aligned to CEFR. A large amount of work has been carried out by testing agencies, language learners and users in their effort to align their language tests to CEFR for over a decade. The collection of work to align language tests against CEFR was compiled and has been turned into a book by Martyniuk (2010) with the aim of highlighting the process of linking a single language test to the CEFR.

CEFR-aligned language tests commonly receive international recognition since they follow international benchmark. Countries such as China, Taiwan, Netherland, Norway and Malaysia have begun to align their tests against the six levels of CEFR descriptors. The process of aligning a test against CEFR could be undertaken using systematic stages as suggested in the manual provided by the Council of Europe. The linking of CEFR to a language test could be conducted from the aspects of familiarization, standardizations, specifications and empirical evidence (Council of Europe 2009).

Linking or comparing scoring rubric against CEFR rating scale is also one of the common practices among test designers who wish to link their language tests to CEFR. CEFR framework is user friendly because of its specific CEFR rating scales of all four English skills namely writing, reading, listening and speaking. CEFR rating scale has been proven as a valid rating scale to be used for rating purposes based on several investigations conducted by Carlsen (2010), Chen (2009), Eckes (2012), Forsberg & Barthing (2010) and Kuiken, Vedder & Gilabert (2010). These studies have shown that the CEFR rating scale is also practical and the descriptors for each level are coherent. CEFR rating scale is also claimed as more rater – friendly, and could promise positive end results after necessary modifications are made to the CEFR rating scale (Harsch & Martin 2012).

Linking or aligning a language test to the CEFR is inevitable because it gives some form of recognition at the international level and at par with other established international language tests such as IELTS, TOEIC and TOEFL. Although the process is complex and the alignment must be made based on the manual provided by Council of Europe, it still ensures the validity and quality of a CEFR aligned language test.

In Malaysia, the implementation of CEFR has involved a major change to the syllabus and assessment to the extent of using totally imported content materials and relying on assessments that are imported from other countries. The implementation of CEFR in Malaysia is thoroughly documented in the latest 2015 – 2025 Malaysia Education Blueprint. According to Hazita (2016), one of the significant elements in the Malaysia Education Blueprint is the alignment of education system against CEFR.

THE STATUS OF CEFR IN MALAYSIA

English is regarded as the language of choice by most organizations and industries in Malaysia. Ever since English has become an essential element for successful business, employers now tend to seek competent users of English to join their organizations. This may sound easy but to employ graduates who are competent and fit the company's requirement is not an easy task. Some employers have shared their frustrations in employing graduates who are competent in English language (Mohd Shukri, Rafikul & Noor Hazilah 2014; Sheridan 2014; Yuen 2015). According to the Ministry of Education's report (2013), Malaysia has produced graduates who were not able to work and compete at the international level due to poor command of English. This is also supported by a report in the mainstream newspaper which claimed poor level of English language proficiency among school and universities graduates despite 12 years of formal English education (The Star 2018). It is an alarming indicator for the Ministry of Education to take a tremendous step to solve this issue, indirectly find new ways to improve the level of English proficiency among graduates.

In 2013, the Ministry of Education has hired Cambridge English to find out the level of English proficiency of all English teachers in Malaysia as a kick start in improving the proficiency of graduates through reformation of the syllabus and assessment. The results showed that majority of English teachers have poor level of English proficiency which is equivalent to CEFR level B2 (The Sun Daily, 2016). Following this, the ministry has decided to review English education, and introduced the reformed English education known as the

English Language Education Roadmap 2015 – 2025 which includes the adoption of CEFR into the curriculum (Abdul Hakim, Radzuwan & Wan Zhafirah 2018). The influence and wide spread of CEFR is inevitable since it is considered as part of globalization which helps to elevate English proficiency level among Malaysian in order to compete economically with other countries (Nurul Farehah & Mohd Salehhuddin, 2018). This is the direct effect of globalisation which leads to the increasing role of English as a lingua franca (Yu, 2019). Malaysia has to participate by ensuring the human resources it produces are skillful and possess good command of English to avoid from being left behind both in the economic and education sectors.

The language roadmap is divided into three phases; phase one was conducted from the year 2013 to 2015. Apart from sending teachers to various trainings, the setting up of CEFR target level for each educational stage was also planned at phase one. The details of target CEFR level set by the Ministry of Education for all educational stages are shown in Table 1.

TABLE 1. CEFR levels against educational stages

Educational stages	CEFR levels
Pre-school	A1
Primary school	A2
Secondary school	B1/ B2
Post- secondary school	B2
University	B2/C1
Teacher education	C1

After CEFR implementation received attention and highlights from academics and public, there have been a number of studies conducted related to CEFR by local researchers in the Malaysian context. Since teachers are the significant figure in the CEFR implementation, many CEFR studies have involved teachers' roles and views towards CEFR implementation. Lo (2018) for instance conducted a study involving over 200 secondary school English teachers to measure the relationship between teachers' concern and anxiety to their acceptance and readiness to implement CEFR in their teaching. This study was mainly conducted based on the application of Concern – Based Adoption Model (CBAM) as the theoretical framework.

Apart from English teachers at school, university academics are also among the respondents of CEFR studies. A study was conducted with the objectives to find out 25 local university English lecturers' views on students' performance in the existing English language courses based on CEFR global scale. This study was also interested to explore and link the content of English language courses to CEFR (Ramiaida, Noor Sazaai, Norhana, Fariza, Zarina & Juliana, 2017). The study signifies universities' interest to link existing English language courses to CEFR in order to improve the content and standard of English courses offered by universities.

There are studies that include students as data source. For instance, Majdah (2018) examined the undergraduates' level of English proficiency according to CEFR levels. Specifically, the study was carried out to gather the levels of English vocabulary and grammar to CEFR global scale and compared the scores of two groups of students in a Malaysian polytechnic. The respondents were 29 peer counsellors and 26 Students Representative Council (SRC) who took a free online CEFR test available on British Council Website. The findings of this study were then used as a guideline to test candidate's level of English proficiency before they were selected as members of Polytechnic peer counsellors and Students Representative Council. There was also another study conducted involving students from a public university. The aim of the study was to identify the English language writing proficiency of 406 undergraduates based on International Islamic University Malaysia (IIUM)'s English Proficiency Test (EPT) and CEFR. The study was also interested to ascertain the relationship between EPT rubric and CEFR global scale. It was found that majority of IIUM undergraduates

were at CEFR level B2 and there was a significant relationship between the IIUM EPT writing scores and CEFR writing scale (Mohd Khairul 2017). Sidhu, Kaur and Lee (2018) conducted a study which investigated the implementation of the CEFR- aligned school-based assessment in primary ESL classrooms. Nevertheless, the findings were not positive and the process can be considered less successful because the teachers were lacked of full understanding of the method. They also admitted they have very limited knowledge and familiarity of the revised CEFR-aligned ESL curriculum.

A similar study with the objective of aligning IIUM EPT against CEFR global scale was conducted by a group of academics of IIUM by Engku Haliza, Isarji, Khairiah, Faridah and Ainon Jariah (2017). The study analyzed 143 writing scripts by grading them based on CEFR global scale and the findings showed strong relationship between CEFR and EPT scoring rubrics within an acceptable one band difference. This study intends to focus on English teachers' awareness and familiarity to CEFR global scale and the changes in the English education system which directly affected them as stakeholders. Previous studies mostly focused on the implementation process of CEFR and the effort to link existing assessment or test to CEFR. There is no study conducted to measure the appropriacy of the current assessment. Hence, it is utmost important to conduct a study which determines the suitability of the current assessment so that the findings could assist Ministry of Education whether to retain or remove SPM English paper assessment items. Indirectly, it would serve as a helpful recommendation to Ministry of Education in regards to the whole implementation process especially the alignment of SPM English paper assessment items against CEFR. Therefore, this study is conducted to address the following objectives:

1. To identify the suitability of the current -writing and reading items against the CEFR level set by the Ministry of Education
2. To find out English teachers' familiarity with CEFR writing and reading scales by suggesting suitable CEFR levels for the items.

METHOD

RESPONDENT

The respondents of this study were 331 high school English teachers. The number of respondents is sufficient as a study on Thai secondary school English teachers' encounters with the CEFR conducted by Franz and Teo (2017) supported the sample size of this study when a total of 323 teachers were involved. The number of respondents of that study which was more or less the same to the present study. The teachers were chosen and found to be suitable respondents because of their 15 to 20 years of teaching experience and they are familiar with the syllabus and format of English national examination papers. The respondents involved in this study were also chosen because they have attended CEFR familiarization workshops organized by Ministry of Education, Malaysia. This is important because respondents who are familiar with CEFR global scale could provide more accurate results compared to those who are not familiar with CEFR global scale and its concepts.

INSTRUMENT

This study uses an assessment checklist to collect the data. The assessment checklist was constructed based on the compilation of SPM English papers from 2012 – 2017 table of specifications. The tables of specifications were followed strictly to ensure the items put

forward in the assessment checklist were the same to the ones that appeared in the exam papers. The tables of specification for SPM English papers from 2012 to 2017 were analyzed carefully to look for recurring patterns or themes. Then, the types of questions which were frequently asked were chosen and included in the assessment checklist.

Three content experts from Universiti Kebangsaan Malaysia (UKM) and Universiti Kuala Lumpur (UniKL) were assigned to verify and validate the assessment checklist. The content expert from UKM has a PhD in Language Testing with more than 30 years of experience as an academic and researcher. Another language expert from UniKL is also a language testing expert with a PhD in Language Testing from University of Bedfordshire, UK with more than 25 years of teaching experience and the second language expert who is a Senior Lecturer at UniKL has a PhD in Linguistics from University of Glasgow with 25 years of teaching and conducting research.

A pilot study was also conducted and SPSS version 23 as well as Winstep version 4.0.1 was used to measure reliability of the assessment checklist. Apart from answering the assessment checklist, the participants were also asked about the construct and face validity of the instrument. All items in the assessment checklist are considered good based on the high value of Cronbach Alpha. Reliability reading of 0.8 and above is considered good and strongly acceptable (Bond & Fox 2007). No necessary amendments were made to improve the checklist since the Cronbach Alpha and reliability readings from both SPSS and Winstep show high values. Table 1 shows the Cronbach Alpha readings of the assessment checklists.

TABLE 2. Reliability statistics of assessment checklists

		Items (n:13)	Teachers (n:30)
SPSS	Cronbach Alpha	.89	-
Winstep	Reliability	.82	.76

PROCEDURES

The English teachers who were involved in this study were recommended by the schools' Heads of English panel. The teachers were briefed about the objectives of the study and their roles as subject matter experts. Then, the assessment checklist along with the SPM English writing and reading papers of 2017 were given to the teachers. The CEFR writing and reading scales were also provided to the teachers as a reference. Clear and specific instructions were given to the teachers to ensure they know how to fill up the assessment checklist. The whole process took approximately 3 weeks to complete.

DATA ANALYSIS

The data of this study were analyzed using two computer software, namely: Winstep and NVivo version 11. Winstep was used to analyze the assessment items and NVivo was used to analyze teachers' comments and feedback. Winstep was used in the data analysis in order to determine the suitability of the assessment items from the value of infit mean square (MnSq). Moreover, Winstep also provides the value of logit and standard error (SE) which indicate items difficulty level. The values produced by Winstep are useful in accomplishing the objectives of this study. Winstep was also chosen because of its ability to analyze persons, items and rating scales. More importantly, Winstep was used in previous studies related to CEFR global scale by Franz and Teo (2017), Zhao, Wang, Coniam and Xie (2017) and Runnels (2013).

Table 2 illustrates the interpretation of infit mean square (MnSq) values and Table 3 shows the interpretation values for both logit and standard error (SE).

TABLE 3. Interpretation values of infit mean square (MnSq)

Recommended value range	Value interpretations	Suggested action
<0.6	Item is too easy	Removed
1.4>	Item is difficult and confusing	Removed
0.6 – 1.4	Item is suitable	Retained

Values interpretation of infit mnsq can be made based on three recommended value range. First, any value which falls below 0.6 is interpreted as items which are too easy which means the items should be removed because they are not suitable. Second, items with the infit mnsq values which are higher than 1.4 are also found to be not suitable since the items are considered difficult and confusing, hence they should be removed. Lastly, only items with the values of infit mnsq which fall within 0.6 to 1.4 could be retained because the values suggest that these items are suitable (Bond & Fox, 2007; Linacre 2014).

TABLE 4. Interpretation values of logit and standard error (SE)

	Values	Interpretation
	High & positive	Difficult item
Logit	Low & positive	Less difficult item
	Negative (-)	Easy item
Standard error (SE)	Big/ High	Difficult item
	Small/ Low	Easy item

Logit value however indicates level of items difficulty and the logit value with high and positive values + mean the items are difficult. The difficulty level of items decreases as the logit values get smaller. Nonetheless, logit values which are negative (-) indicate that these items are easy. The level of difficulty measures through logit values are supported by the S.E values although the interpretation of the S.E values differ from logit values. Items with small S.E value would have a high logit which shows confidence in measure. In a group of items, high S.E value indicates that the items are easy and the level of difficulty increases as the S.E value gets smaller.

Teachers' comments and feedback were typed and transferred into the form of word document, then the comments and feedback were sorted out and coded according to writing items and reading items. The process of data coding is to assist qualitative content analysis suggested by Krippendorff (2013). Then, the raw data were keyed-in into NVivo version 11 to be processed to look for salient findings and recurring themes.

FINDINGS AND DISCUSSION

TABLE 5. Assessment items measurement report

Assessment items	Logit	Standard error (S.E)	Infit Mean Square
Open topic	-.84	.24	1.09
Argumentative	-.73	.24	.82
Novel	-.62	.24	.82
Descriptive	-.56	.23	.86
Talk	-.23	.23	.83
Summary	-.18	.23	1.29
Letter	-.02	.23	1.23
Reading comprehension	.03	.23	1.04
Poem	.35	.23	.89
Extract	.46	.23	.69
Advertisement	.57	.23	1.32
Information transfer	.72	.23	.92
Notice	1.04	.23	1.11

Table 5 shows the details of the assessment items measurement report which consist of logit, standard error (S.E) and infit mean square (mnsq) values. The results show that there are 6 assessment items with positive logit values and seven assessment items with negative logit values. It can be seen clearly that the three easiest assessment items are open topic, argumentative and novel. It is because open topic, argumentative and novel have negative logit values of -.84, -.73 and -.62 each. This is supported with the S.E values of these assessment items which are 24. The S.E value of 24 is the biggest value which means the items are supposed to be easy. Apart from open topic, argumentative and novel which have negative logit values, descriptive, talk, summary and letter also appeared to have negative logit values. The logit values for descriptive, talk, summary and letter are as follow: -.56, -.23, -.18 and -.02. These assessment items are considered as easy items because of the negative logit values, nevertheless these items are slightly difficult and not as easy as open topic, argumentative and novel because the S.E value is 23. An increase with the S.E value indicates an increase of item difficulty level.

The remaining six assessment items, reading comprehension, poem, extract, advertisement, information transfer and notice share the same S.E value of 23. Therefore, the difficulty level of these assessment items is determined and differentiated based on the logit values. The most difficult assessment item is notice with the biggest logit value of 1.04. The teachers agreed that answering questions related to notice is the most difficult among students. Information transfer, advertisement and extract are the assessment items which are less difficult compared to notice because the logit values get smaller. Hence, information transfer can be considered as the second most difficult assessment items with logit value of .72. This is followed by advertisement with logit value of .57, extract with logit value of .46, poem with logit value of .35 and reading comprehension with logit value of .03. The results conclude that only reading assessment items have positive logit values and considered as difficult items except novel and all writing assessment items were found to be the easy items with the negative logit values and S.E value of 24.

The suitability of the assessment items is determined based on the infit mnsq values. The infit mnsq values of these assessment items fall within the recommended range of 0.6 to 1.4 which mean all these assessment items are suitable to be tested on Form 5 students in English SPM papers 1 and 2. The results also indicate that both easy and difficult assessments items were found to be suitable and needed in testing students' writing and reading skills abilities. The infit mnsq values for open topic, argumentative, novel, descriptive, talk, summary, letter, reading comprehension, poem, extract, advertisement, information transfer and notice are 1.09, .82, .82, .86, .83, 1.29, 1.23, 1.04, .89, .69, 1.32, .92 and 1.11 each. The teachers also used their judgements as subject matter experts to suggest suitable CEFR levels for the assessment items and the findings are illustrated in Table 6.

TABLE 6. Suggested CEFR levels

Assessment items	CEFR level
Open topic	C1
Argumentative	C1
Novel	C1
Descriptive	C1
Talk	B2
Summary	B2
Letter	B2
Reading comprehension	B2
Notice	B2
Extract	B2
Advertisement	B2
Information transfer	B2
Poem	B1

Clearly, most of the assessment items were suggested with CEFR level B2. There is only one assessment item with CEFR level B1 and CEFR level C1 for the remaining four assessment items. Descriptive, talk, summary, letter, reading comprehension, notice, extract, advertisement and information transfer were suggested to be at CEFR level B2 according to the majority of the teachers. Only poem is suggested with CEFR level B1. The suggested levels recommended by the teachers to descriptive, talk, summary, letter, reading comprehension, notice, extract, advertisement, information transfer and poem are found to be suitable because they match the logit values as well as the S.E values. These eleven assessment items have positive logit values under the value of 1.00 which means the level of difficulty is intermediate. Directly, it means CEFR level B2 suggested by the teachers are suitable for these assessment items. Poem has the second lowest logit value of .35 and this indicates that poem is the least difficult item compared to extract, advertisement, information transfer and notice. Hence, the CEFR level B1 as suggested by the teachers' matches the logit value and thus making it suitable.

In contrast, CEFR level C1 suggested by the teachers for assessment items such as open topic, argumentative, novel and descriptive were found to be not suitable because they do not match the logit values and S.E values of these items. These assessment items appeared to be the four assessment items with the lowest logit values and a bigger S.E value compared to other assessment items. The logit values and S.E values of assessment items open topic, argumentative, novel and descriptive indicated that these items are among the easiest items, therefore suggested CEFR level C1 is not appropriate as it does not match the logit and S.E values. The more appropriate CEFR level for assessment items open topic, argumentative, novel and descriptive based on the logit and S.E values could be B1 or A2 because CEFR level C1 is too high and difficult for easy items.

Apart from suggesting suitable CEFR level for the assessment items, the teachers also left their comments and feedback on the items. Table 7 shows the salient comments for writing assessment items.

TABLE 7. Writing assessment items

Writing items	Teachers' comments and feedback
Talk	1. Help to prepare for real life.
	2. A good skill to be practiced in real life situation.
Argumentative	1. Questions are related to real life situations.
	2. A skill that is required in debating issue/ problems in real life situation.
Descriptive	1. Applicable in students' future working field.
	2. Free from question. Easier for students to channel their creativity.
	3. It depends on how the candidates attempt the question which includes the use of precise vocab, writing skill and variation of sentence structure.
Open topic	1. May apply to average and above level of proficiency students.
	2. Questions are related to facts and it will be challenging for intermediate students.
	3. Questions are too vague. Students will face difficulties in answering these questions.
	4. A suitable task in eliciting students' personal views on a situation / subject/ topic.
Summary	1. It doesn't really match with CEFR descriptors and not emphasized in lower secondary level.
	2. Train students to identify main points.
	3. Train students to select specific information.
	4. Points provided to help candidates.
	5. C2 – able to rephrase, C1 – able to use short well sentences, B2 – able to identify all the points.
Letter	1. Irrelevant to current situation (informal letter). Formal letter is still relevant to present situation.
	2. Questions are related to real life situations and students can relate to themselves.
	3. Students' works are marked according to the format.
	4. Students can express themselves.
	5. Suggests informal letter to be changed to email. Formal letter should be retained.
	6. Adjust the rubrics to match the CEFR descriptors for each band. Topics are varied and suitable for all bands.
	7. The rubrics of the question can be worded to suit the target level. More or less notes can be given to differentiate the level of guidance given.

Writing assessment items talk, argumentative and descriptive received positive feedback by the teachers. Talk, descriptive and argumentative writing were also highly recommended by the teachers to be retained and to be used with the new CEFR aligned Form 5 English syllabus in 2021. The justification to retain these items are simple; these items are related to students and applicable in real life situation. A few comments left by the teachers were *"Help to prepare for real life"*, *"A skill that is required in debating issue/ problems in real life situation"* and *"Applicable in students' future working field"*.

Writing formal or informal letter is the first task for writing assessment item (English SPM paper 1), thus it received the most comments from participants. Among the salient comments are to omit informal letter because the participants viewed it as irrelevant to the current situation. It was even suggested that informal letter should be replaced with email. Nonetheless, most of the teachers agreed that formal letter is still relevant and therefore shall be retained as the students will use it in real life.

Some instances of the comments are *"Irrelevant to current situation (informal letter). Formal letter is still relevant to present situation"* and *"Suggests informal letter to be changed to email. Formal letter should be retained"*. Suggestions by teachers to replace informal letter with email is logic and make sense considering email is widely used as medium of communication in present days. Moreover, email is preferred by many organizations particularly when there is a need for internal communication in an organization. Indirectly, the use of email instead of informal letter prepares our high school leavers before they venture into the real working world. This also promotes meaningful learning to school students since the knowledge they gain in school can be used outside the school.

In addition to that, the teachers also suggested rubric adjustment for letter in order to align this assessment item against six levels of CEFR descriptors. The teachers stated that *"Adjust the rubrics to match the CEFR descriptors for each band. Topics are varied and suitable for all bands"*. The teachers also wrote similar suggestion about rubrics *"The rubrics of the question can be worded to suit the target level. More or less notes can be given to differentiate the level of guidance given"*. This suggestion implies that the teachers were aware of the content of six CEFR level descriptors and they knew that a letter could be written and assessed differently depending on the CEFR level required.

In contrast, open topic is one of the assessment items which received unpleasant comments from the teachers. They believed that open topic is no longer applicable and also inappropriate to measure students' ability in writing. The teachers were quoted stating *"May apply to average and above level of proficiency students"*. They also claimed that open topic is not suitable for high school students because *"Questions are related to facts and it will be challenging for intermediate students"*. The teachers also mentioned that open topic is irrelevant because of the structure of the item itself. They wrote *"Questions are too vague. Students will face difficulties in answering these questions"*.

Two main different views were identified for summary. The teachers claimed that summary is not applicable since the exposure given to the students is rather late. It is because students started to learn about summary when they reach upper secondary and this is only to prepare them for SPM. According to them *"it doesn't really match with CEFR descriptors and not emphasized in lower secondary level"*. They also claimed that summary does not really match CEFR global scale but no justification and suggestion was provided as solutions. On the other end, teachers gave positive remarks by providing specific comments as to how summary could be aligned against CEFR descriptors. The comment left was *"C2 – able to rephrase, C1 – able to use short well sentences, B2 – able to identify all the points"*. The recommendation from teachers indicates that summary needs to be amended to ensure it matches with CEFR descriptors. It also shows that different level of CEFR descriptors would cater different ability of learners.

In brief, it is found that only four items were recommended to be retained and the recommendation includes rubrics adjustment so that these items match CEFR descriptors well. One item is to be omitted but no suggestion was provided as replacement. Similar to writing assessment items, the respondents' written comments and feedback for reading assessment items are presented in Table 8.

TABLE 8. Reading assessment items

Reading items	Teachers' comments and feedback
Notice	1. Students are aware of such problems.
	1. It relates to the real world.
Advertisement	2. Stimuli must be clear, easy and precise so begin with A1, A2 would be able to answer.
Information transfer	1. It extracts information from reading.
	2. Students must have high proficiency level to the information correctly.
Reading comprehension	1. Students must master how to select main points to answer 5WH questions and to answer questions.
	2. We need a lot of reading in real life.
Extract	1. Depends on the level of lexical items used in the text rather than the text type.
	2. It can be done in comprehension section.
	1. It promotes reading.
	2. Students analyze the novel according to specifications.
Novel	3. Students required to express their own ideas.
	4. Depending on the target answer. From the answer given, we will be able to place the level.
	1. It promotes analytical skill.
	2. Train students to think underlying messages in a poem.
Poem	3. Train students to analyze connotative meaning and messages conveyed by the writer.
	4. Easy to attempt

Table 8 summarizes teachers' comments and feedback for reading assessment items. The two salient themes of these comments are comments related to CEFR global scale and general structure of assessment items. First, the teachers believed some of these assessment items are matched to CEFR and its descriptors. For instance, the teachers agreed that advertisement used to assess students reading skills is suitable and matches CEFR global descriptors. The teachers also stated in specific CEFR level that would suit this item. As extracted from the checklist, *"stimuli must be clear, easy and precise so begin with A1, A2 would be able to answer"*. Similarly, the teachers also emphasized on altering the literature section in particular the novel. It was highlighted that necessary alteration is needed so that novel and questions related to it can be aligned against CEFR. The teachers were quoted *"depending on the target answer. From the answer given, we will be able to place the level"*.

Aside from providing comments based on CEFR descriptors, some of the teachers also stated several positive comments about these items specifically poem and novel. Most of the comments show that these items should be retained although new form 5 English CEFR aligned syllabus will be used soon. Some of the comments are "*it promotes analytical skill*", "*train students to identify main points*" and "*students required to express their own ideas*". Nonetheless, the teachers were in agreement that reading assessment items of information transfer and reading comprehension were challenging for the students. The teachers provided several reasons in which they believed reading assessment items of information transfer and reading comprehension could be removed. The teachers agreed that *"Students must have high proficiency level to the information correctly"* in order for them to answer questions related to information transfer. Reading assessment item of reading comprehension was also viewed negatively by the teachers because they agreed that *"Students must master how to select main points to answer 5WH questions and to answer questions"* and the students *"Need a lot of reading in real life"*. In conclusion, teachers' positive comments support the value of infit mnsq that all reading assessment items are suitable and shall be retained. Teachers' evaluations as subject matter experts using assessment checklists should be taken into consideration.

CONCLUSION

In conclusion, all the writing and reading assessment items are suitable with a combination of three easy items, six intermediate items and four difficult items. Some of the writing and reading assessment items are found to be relevant against CEFR global scale as well as the new CEFR aligned English syllabus. Therefore, the suitable and relevant items should be retained instead of to be removed. These assessment items can be carefully aligned to CEFR global scale. The findings of present study are also supported by O'Dwyer (2014) who claimed that CEFR implementation does not necessarily involve total revamp of the current syllabus and assessment so that it can be replaced with the new CEFR aligned syllabus and assessment. CEFR can be a common language framework for stakeholders to evaluate existing practices (Papageorgiou 2014); then it would be sufficient to evaluate the current English reading and writing assessment items and aligned it to CEFR wherever necessary instead of totally abolishing the current assessment items.

The Ministry of Education has set CEFR levels B1/B2 as the target level for upper secondary school. The teachers proved that they were aware of the target level set by the ministry when they suggested that CEFR level that matches the target CEFR level. The findings also revealed that the teachers were familiar with the descriptors of CEFR reading and writing scales because they were able to match it accordingly. They have also proved that their judgements as subject matter experts could be used in deciding the appropriate CEFR level for the assessment items. Their judgments are also supported with the logit as well as S.E values. Nonetheless, the teachers had problems with consistency in judging and recommending suitable CEFR level of the writing and reading assessment items when they suggested four CEFR levels which were not suitable and did not match the logit values and S.E values.

The teachers as the subject matter experts suggested the appropriate CEFR level based on their expertise and knowledge as senior year teachers and national examination assessors against the six levels of CEFR descriptors but it does mean they are able to accurately match the assessment items with CEFR global scale since they were not CEFR experts. This explains such results particularly when the teachers overrated some easy assessment items with CEFR level C1 instead of a more suitable level such as CEFR level A2. This supports the argument put forward by the Council of Europe (2009) which mentioned linking a qualitatively poor examination to the CEFR is a waste enterprise. Nonetheless, it is not fair to claim SPM English paper as a poor examination and should not be aligned to CEFR global scale; moreover, the teachers only suggest the suitable CEFR level for the assessment items within their capacity as subject matter experts.

The CEFR level suggested by teachers for assessment items as subject matter experts can be used by the MOE by comparing teachers' suggestions with suggestions made by Cambridge English as CEFR experts. The combination of subject matter experts and CEFR experts would result in the more suitable CEFR level set for English syllabus and assessment. Other than that, this study also implies the importance of including teachers and their expertise as well as their voices in the implementation from the beginning.

A total of 331 teachers who were involved in this study however only represented a small number of English teachers in Malaysia. For this reason, future study is highly recommended to get a larger sample size by involving more teachers since the present study survey English teachers in Selangor, Kuala Lumpur and Putrajaya only. Therefore, the findings of the future study could be improved and perhaps could be more accurate when more teachers from other states particularly those in rural areas are included throughout Malaysia. It is also highly recommended that the new CEFR aligned assessment which could be used at all levels of educational stage to be studied because it is crucial to measure the significant change and the impact on teaching and learning as well as the quality of English proficiency among

students. Moreover, the findings of studying the new CEFR aligned assessment could reveal if it is really necessary for Ministry of Education to make a total revamp of the current English syllabus as well as assessment and replace it with the new CEFR aligned English syllabus and assessment.

REFERENCES

Abdul Hakim Ali Abdul Aziz, Radzuwan Ab Rashid & Wan Zhafirah Wan Zainudin. (2018). The enactment of the Malaysian common European framework of reference (CEFR): National master trainer's reflection. *Indonesian Journal of Applied Linguistics. Vol. 8(2),* 409 – 417. doi:10.17509/ijal.v8i2.13307

Bond, T.G & Fox, C.M. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. (2nd) Ed. New Jersey: Lawrence Earlbaum Associates Incorporation.

Byram, M., Gribkova, B & Starkey, H. (2002). *Developing the Intercultural Dimension in Language Teaching: A practical introduction for teachers*. Language Policy Division. Directorate of School, Out of School and Higher Education. Strasbourg: Council of Europe

Carlsen, S. (2010). Discourse connective across CEFR – levels: A corpus based study in I, Martin, M and Vedder, I (Eds.). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp.81 – 99). EUROSLA Monoseries 1.

Chen, Y.H. (2009). Investigating lexical bundle across learner writing development. Published PhD Thesis, Lancaster University, Lancaster, United Kingdom.

Council of Europe. (2009). Relating language examinations to the Common European Framework of Reference for Languages. Learning, Teaching, Assessment. CEFR: A manual. Language Policy Division, Strasbourg.

Eckes, T. (2012). Examinee – centered standard setting for large – scale assessments: The prototype group method. *Psychological test and assessment modeling Journal. Vol.54(3*), 257 – 283. Retrieved September 26, 2018 from https://www.testdaf.de/fileadmin/Redakteur/PDF/Forschung-Publikationen/Eckes_PTAM_2012-3.pdf

Engku Haliza Engku Ibrahim, Isarji Sarudin, Khairiah Othman, Faridah Abdul Malik & Ainon Jariah Muhammad. (2017). The assessment of writing within the CEFR scale: A Malaysian context. *Advanced Science Letters Journal. Vol. 5(3)*, 4944 – 4947. doi: 10.1166/asl.2017.8986

Forsberg, F & Bartning, I. (2010). Can linguistics features discriminate between the communicative CEFR levels? A pilot study of written L2 French in Barthing, I, Martin, M and Vedder, I (Eds.). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. (pp.81 – 99). EUROSLA Monoseries 1.

Franz, J & Teo, A. (2017). "A2 is normal" – Thai secondary school English teachers' encounters with the CEFR. *RELC Journal. Vol. 49(3),* 1 – 7. doi: 10.1177/0033688217738816Harsch, C & Martin, G. (2012). Adapting CEFR descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing writing. Vol. 17(4)*, 228 – 250. Retrieved September 29, 2018 from https://www.sciencedirect.com/science/article/abs/pii/S1075293512000244

Hazita Azman. (2016). Implementation and challenges of English Language Education Reform in Malaysian Primary Schools. *3L: The Southeast Asian Journal of English Language Studies. Vol. 22(3).* 65 – 78. doi: 10. 17576/3L-2016-2203-05

Krippendorff, K. 2013. *Content Analysis: An Introduction to its Methodology*. Los Angeles: Sage Publication.

Kuiken, F, Vedder, I & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing in Barthing, I, Martin, M and Vedder, I (Eds.). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research.* (pp. 81 – 99). EUROSLA Monoseries 1.

Linacre, J. M. 2014. A user's guide to FACETS Rasch-model computer programs. Chicago, IL. Retrieved from http://www.winsteps.com/a/facets-manual.pdf [14 October 2017].

Little, D. (2011). The Common European Framework of Reference for Languages: A research agenda. *Language Teaching Journal. Vol. 44(3)*, 381- 393. doi: 10.1017/s0261444811000097

Lo, Y.Y. (2018). English teachers' concern on Common European Framework of Reference for Languages (CEFR): An application of CBAM. *Jurnal Kurikulum dan Pengajaran Asia Pasifik. Vol. 6(1),* 46 -58. Retrieved September 2, 2018 from http://juku.um.edu.my/index.php/JUKU/article/view/11174/7649

Majdah Mahamud. (2018). The CEFR levels of English among student representatives in a polytechnic in Malaysia. National Innovation and Invention Competition through Exhibition (iCompEx). 1 – 10. Retrieved September 5, 2018 from https://upikpolimas.edu.my/conference/index.php/icompex/icompex_2018/paper/viewFile/382/247

Martyniuk, W. (2010). *Aligning tests with the CEFR: Reflections on using the Council of Europe's draft manual*. Cambridge. Cambridge University Press.

Mohd Khairul Abu Sufi. (2017). Mapping the IIUM students' English language writing proficiency to the Common European Framework of Reference for Languages (CEFR). Published Master's Thesis. Kuliyyah of Islamic Revealed Knowledge and Human Sciences. International Islamic University, Gombak, Malaysia. Retrieved September, 2018 https://lib.iium.edu.my/mom/services/mom/document/getFile/cOdHBXR3nHvajYPmrc2GCHUsHLqe YCgw20171113173437936

Mohd Syukri Abdul Hamid, Rafikul Islam & Noor Hazilah Abd Manaf. (2014). Malaysian's graduates employability skills enhancement: An application of the importance performance analysis. *Global Business Advancement Journal*. Vol.7 (3). 181 – 197. Retrieved January 15, 2016 from https://rafikulislam.com/uploads/myworks/43167429855ceb3340cc14.pdf

Nurul Farehah Mohamad Uri & Mohd Salehhuddin Abd Aziz. (2018). Implementation of CEFR in Malaysia: Teachers' awareness and challenges. *3L: The Southeast Asian Journal of English Language Studies*. *Vol. 24 (3),* 168 – 183. doi: 10.17576/3L- 2018-2403-13

O'Dwyer, F. (2014). Towards critical, constructive assessments of CEFR – based language teaching in Japan and beyond. *Osaka University Knowledge Archive*. Vol. 31(3). 191 – 204. Retrieved September 11, 2018 from https://ir.library.osaka-u.ac.jp/repo/ouka/all/51427/slc_41-191.pdf

Papageorgiou, S. (2014). Issues in aligning assessments with the Common European Framework of Reference. *Language Value* .*Vol. 6(1),* 15-27. doi:10.6035/languageV2014.6.3

Ramiaida Darmi, Noor Saazai Mat Saad, Norhana Abdullah, Fariza Puteh Behak, Zarina Ashikin Zakaria & Juliana Niza Ismail Adnan. (2017). Teachers' views on students' performance in English language proficiency courses via CEFR descriptors. *IJAEDU- International E-Journal of Advances in Education*. *Vol. 3(8),* 363 – 370. Retrieved September 1, 2018 from http://ijaedu.ocerintjournals.org/download/article-file/338673

Runnels, J. 2013. Student ability, self – assessment and teacher assessment on the CEFR – J's can do statements. *The Language Teacher Journal*. *JALT SIG Special Issue. Vol. 3(5),* 3 – 5. Retrieved from https://jalt-publications.org/files/pdf-article/37.5tlt_art01.pdf [16 September 2016]

Sheridan, M. (2014). Job seekers with A in SPM English but can't speak a word of it. The Malaysian Insider. 26 June: 20.

Sidhu, G.K., Kaur, S & Lee, J.C. (2018). CEFR-aligned school based assessment in the Malaysian primary ESL classroom. *Indonesian Journal of Applied Linguistics. Vol. 8(2)*, 452- 463. doi:10.17509/ijal.v8i2.13311.

The Ministry of Education Report. (2013). Malaysia Education Policy Review: Abridged Report. May 2013. Retrieved from http://unesdoc.unesco.org/images/0022/002211/221132e.pdf

The Star Online. (2018). Let's have our own textbooks.4 February. https://www.thestar.com.my/news/education/2018/02/04/lets-have-our-own-textbooks/

Yu, H.L. (2019). Rethinking globalization, English and Multilingualism in Thailand: A report on a five year ethnography. *3L: The Southeast Asian Journal of English Language Studies. Vol. 25(1)*, 69 – 84.

Yuen, M. (2015). Poor English a major handicap. The Star. 15 November: 23.

Zhao, W., Wang, B., Coniam, D & Xie, B. 2017. Calibrating the CEFR against the China Standard of English for College English vocabulary education in China. *Language Testing in Asia Journal. Vol.* 7(5), doi: 10.1186/s40468-017-0036-1