Aligning a University English Language Proficiency Measurement Tool with the CEFR: A Case in Malaysia

Nurul Najwa Baharum*, Lilliati Ismail, Nooreen Nordin and Abu Bakar Razali

Department of Language & Humanities Education, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

The Common European Framework of Reference (CEFR) describes the capability of learners' language skills at six reference levels. It is internationally recognised as the standard language proficiency framework for describing language learning, teaching and assessment. Many countries, including Malaysia, have attempted and invested tremendous efforts to adopt the CEFR as a reference for language ability at all levels of education. However, there are many ways of adopting CEFR, and it is a continuous process of alignment between curriculum and assessment. In this regard, this study is carried out to examine how a Malaysian university attempts to demonstrate this alignment by correlating the scores obtained from English language proficiency courses in the university, called the English Language Competence Score Average (ELCSA), to a CEFR-aligned English language proficiency test (Linguaskill). The results showed an overall significant positive correlation that varied in strength. The overall correlation was 0.371, a positive but weak correlation whereby the strongest correlation was seen between ELCSA and CEFR Writing score with a correlation of 0.417, which is positive and moderate in strength. Therefore, it could be identified that a score of 3.25 and 3.5 on the ELCSA can be considered equivalent to a Linguaskill score of 160 (CEFR Band B2). It could be considered that the B2 CEFR level could be subdivided into lower and higher B2. However, there is a need to correlate ELCSA with other CEFR-aligned tests and perform further revisions to the English language proficiency programme at the university to successfully benchmark the programme and its assessment tool, ELCSA, with the CEFR.

Keywords: Assessment, benchmarking, CEFR, English language programme, language testing

INTRODUCTION

There are increasing concerns in establishing standards for the English Language in terms of international benchmarking worldwide (Read, 2019); such language benchmarks standard can be an expressive scale of language ability (Inguva, 2018). Establishing these standards can be quite important in securing places in international higher education institutions and for employment in international companies. The Common European Framework of Reference, CEFR, has become an international benchmark for language competency in many countries, even beyond Europe, such as Mexico, Canada, Japan, and Vietnam. Additionally, many international high-stakes tests such as the IELTS, TOEFL and TOEIC have now been aligned to the CEFR, further underscoring how the framework has gained acceptance and credibility worldwide (Don & Abdullah, 2019). However, despite CEFR being adopted worldwide, research has also claimed that the CEFR still lacks links with stakeholders, socio-educational contexts and empirical validation (Ali et al., 2018).

The Malaysian government acknowledges and stresses the mastery of the English language to gain economic and social leverage in the globalised world. Therefore, it is essential to establish standards and benchmarks that are accepted worldwide to measure proficiency levels among Malaysians. The English Language Standards and Quality Council (ELSQC) and the English Language Teaching Centre (ELTC) of the Malaysian Ministry of Education were given the task to align the Malaysian English education curricula and assessment with the CEFR, as well as to develop a roadmap for systematic reform of Malaysia's English language education (Prakash, 2019). Following this educational shift, the primary and secondary levels of education have replaced their English language textbooks with CEFR-aligned textbooks. These actions were also followed by the alignment of SPM and MUET examinations in which the results of the test takers English language proficiency were banded against the CEFR descriptors (Sufi & Stapa, 2020).

One of the key issues that surfaced during the adoption of the CEFR in Malaysia was the fear that the Malaysian National Education Philosophy would be side-lined and European cultural values and elements would instead dominate local and national content (New Strait Times, 2019). However, the ministry has organised programmes for teacher training, curriculum familiarisation and adaptation, as well as continued efforts in providing more resources. The use of of-the-shelf CEFR-aligned textbooks (as textbooks for National primary and secondary schools) that were carefully selected, vetted and revised to suit the Malaysian context, by working closely with the publishers, has proven to be more cost-effective and offers a wider acceptance of other cultures along with providing a variety of ways of using the English language in different contexts (Sani, 2018). Implementing the English Language Education Reform in Malaysia was foreseen to be complex, costly and requires persistent efforts and tremendous patience. However,

all these endeavours are for the national advancements that will benefit Malaysia's current and future generation (Ministry of Education, 2015).

Research Problem

While there are some positive indications in referencing language performance against the CEFR at the pre-tertiary level, as stated in the Cambridge Evaluation study in 2017, the impact of such an initiative is not yet seen at the university level (Zulkefli, 2017). The English Language Education Reform in Malaysia stresses the importance of implementing the CEFR in universities to address problems related to poor English communication skills among graduates, which would inadvertently negatively affect their learning experience, employability potential and realise the national agenda (Ministry of Education, 2015). Based on The Roadmap, it was stated that university students are to possess a CEFR B1 level upon university entrance. The Department of Higher Education, Malaysia, stipulated that 'international students' must also sit for exams that reference the CEFR to fulfil the English requirements for university admission purposes (Jaafar, 2019). Additionally, students are required to reach a proficiency of CEFR B2/C1 upon graduation. In accomplishing the required CEFR condition among university graduates, The Roadmap implies that students' English language proficiency may need to be reassessed by the institution prior to their completion of studies (Sufi & Stapa, 2020).

Furthermore, the Malaysian Qualifications Agency (MQA), responsible for quality assurance and accreditation of Malaysian universities, specified that English proficiency courses that are not equated with the international benchmark, CEFR, can no longer be used to fulfil university requirements (Malaysian Qualification Agency, 2020a). Furthermore, the MQA also stated that a CEFR minimum proficiency of C1, or its equivalent in the relevant language, is required to pursue certain job positions in tertiary level institutions (Malaysian Qualification Agency, 2020b). Therefore, there are increasing attempts, demands, and a heightened level of importance for universities to be CEFR-aligned not only of their courses but also their entry and exit grade requirements of universities.

Noticeably, the alignment of CEFR within the tertiary level of education is underexplored and is an area of concern in which further research is required as it affects the efficiency of the English Language proficiency among university graduates. Given that the CEFR is required by the Ministry of Education Malaysia to be aligned with the curricula in the tertiary level education and considering that the Malaysian government has invested a substantial amount of money and effort in aligning the curriculum, it is important to investigate the alignment of English language proficiency assessments at different levels of education to the CEFR standards. Furthermore, investigating the alignment between university English language proficiency evaluation measures

and the CEFR could provide invaluable information to policymakers and test developers about the predictability and comparative values of the university English proficiency assessment with a well-recognised international standard for language education and assessment, namely the CEFR.

Hence, this paper seeks to fill the gap in the literature by attempting to align the accumulative scores obtained by undergraduate students who took an English proficiency programme in a Malaysian public university with the CEFR scores based on the Linguaskill test students sat for. The English proficiency score selected for this study is the ELCSA accumulative score. ELCSA stands for English Language Competence Score Average, an accumulated score derived from a package of English language proficiency courses in University Putra Malaysia. More specifically, the paper will attempt to firstly examine the relationship between the scores obtained in the ELCSA and the overall as well as individual language skill scores on the Linguaskill test, and secondly, identify the ELCSA score that is equivalent to a CEFR B2 level which has been targeted as the minimum CEFR level for Malaysian university graduates. By doing so, the paper can contribute to a greater understanding and contextualisation of the CEFR. Furthermore, aligning ECLSA scores to the CEFR will help provide comparative scores in ELCSA with Linguaskill. It could then provide indications of test-takers CEFR levels based on ELCSA accumulative proficiency scores and could assist and contribute to the university's benchmarking efforts of an internally developed English proficiency measurement tool with international standards.

LITERATURE REVIEW

Benchmarking

There are various definitions given on the concept of benchmarking in the literature. For example, Bogan and English (1994) stated that benchmarking is the continuous pursuit of best practices. By establishing measurement points, comparisons can be made for reasons of learning, adapting and ultimately resulting in better performance, which is the main purpose of benchmarking (Fisher, 1996). The essence of benchmarking is also inspiring ongoing learning and boosting organisations to be at their best (Zairi, 1996). The intent of benchmarking is to aid organisations in establishing a baseline performance criterion that should be complied with (Nwabuko et al., 2020). Similarly, Keegan and O'Kelly (2012) consider benchmarking as a method of comparison between organisations to obtain insights from each other. Benchmarking is operative in identifying best practices, and these practices are applied for the benefit of the organisation (Alosani et al., 2016).

In education, especially with second language learning, benchmarking is required when measurable standards are set for learning (Inguva, 2018). Benchmarking in assessments ordinarily attends to the purposes of evaluating and monitoring program efficiency, planning curriculum and

instruction, communicating expectations for learning and predicting future performance whereby it would operate finest when it is specifically designed to deliver the data required for enhancements to be made (Herman & Osmundson, 2010). Benchmarking could also provide information about the position of a specific student, class, or institution in terms of ranking (Canadian Language Benchmarks, 2012). There are many views to the term benchmarking as it is used in various contexts. Nevertheless, it could be said that benchmarking is an ongoing process of seeking the best practices by making comparisons and creating points of reference so that the effectiveness of a particular programme could be identified and further improved.

Regarding benchmarking language learning and assessment, the CEFR framework has proven to be an influential baseline for the development of language curricula and assessment around the world (Read, 2019). However, benchmarking curricula to the CEFR has brought a great deal of discussion whereby some countries found it problematic to strike a balance between the appeal of establishing mutual international standards and the importance of representing the unique educational and social contexts of distinct countries in language learning (Read, 2019). In Taiwan, for example, attempts were made to adopt the CEFR, which meant that their recognised tests needed to be calibrated against the CEFR (Wu, 2012). However, Wu (2012) pointed out that there were several problems

with the process of calibrating tests to the framework, such as the conceptual difficulty in comparing the results of tests that have been designed differently and the lack of technical expertise to confirm the alignment of CEFR upon their tests. Furthermore, Wu (2012) mentioned an unclear relationship between the assessment of English language proficiency according to the CEFR and the grading criteria used by universities. Additionally, it was reported that Taiwan students did not have the exposure to the language to use it communicatively as described on the CEFR scales (Cheung, 2012).

On the other hand, there were also instances where some researchers suggested developing a new framework of reference altogether. For example, in China, rather than adapting the CEFR, the development of a Common Chinese Framework of Reference for Languages (CCFR) or currently known as China Standards of English (CSE) which has been established without much reference to the other frameworks and with their separate tests as measures of student achievement was proposed (Jin et al., 2017). Meanwhile, there are instances where these efforts to align the CEFR with curricula succeeded. For example, in Japan, a team of language researchers at the Tokyo University of Foreign Studies undertook a project to adapt the CEFR to the Japanese context, which successfully resulted in a version of the framework labelled CEFR-J whereby they added sublevels (A1.1, A1.2 and A1.3) to reflect better the degree of English ability (Markel, 2018).

English Proficiency Courses in a Malaysian University

One of the measures taken by universities to improve English language proficiency among students is to offer a range of English language courses required for students to pass as part of their graduation requirements (Rethinasamy & Chuah, 2011). It is also a measure taken by one of the research universities in Malaysia, Universiti Putra Malaysia (UPM). They revamped their English language proficiency level courses and developed an innovative package referred to as the English Language Experience (ELEx). The purpose of the ELEx package is to engage students with the language in a variety of formal and informal situations as well as involve more student-centred courses and task-based language activities. ELEx consists of three components, namely conventional courses (LPE), non-alphabet grade preparation courses (CEL), and language activities (LAX). The number of English skills courses, CEL courses, and LAX activities that students need to take is determined by the MUET results obtained before students start their studies at UPM. Therefore, students who get low results in MUET need to take more English courses and activities than students who achieve high results.

UPM enhanced the ELEx package by implementing a cumulative, and summative assessment of language performance referred to as the English Language Score Competency Average (ELCSA). The ELCSA is obtained calculating average achievement points for the two components of ELEx, namely conventional courses (LPE) and preparatory courses (CEL). LAX activities are not included in this calculation because they serve as support (scaffold) to forming English language skills by allowing students to use the language and build confidence in its use. The assessment of this English language achievement, named English Language Competence Score Average (ELCSA), will be calculated at the final stage of the study program and will be stated in the student transcript. It is also important to add that the ELSCA is isolated from the existing CGPA. Therefore, it does not interfere with nor affects the student's CGPA. Therefore, ELCSA serves as a cumulative summary of the student's achievement in their English language skills.

As mentioned, the targeted level for university graduates is the B2 level of the CEFR whereby at this level, it is expected that graduates can understand complex texts, tackle other abstract topics, engage in discussions as well as be able to communicate with native speakers with ease (Ministry of Education, 2015). A major motivation in introducing the ELCSA is to provide a measure that can indicate the student's English language performance according to the CEFR bands. The ELSCA scores could act as a comparison point compared to other CEFR achievement tests such as IELTS, TOEFL, Linguaskill and MUET. In addition, it could evaluate the effectiveness of the ELEx package.

The ELEx Program at UPM and the Development of ELCSA

In 2013, the Centre for the Advancement of Language Competence (CALC) in UPM executed its undergraduates' English Language Experience (ELEx). The students must follow a carefully developed set of courses and activities during the whole duration of their educational programme.
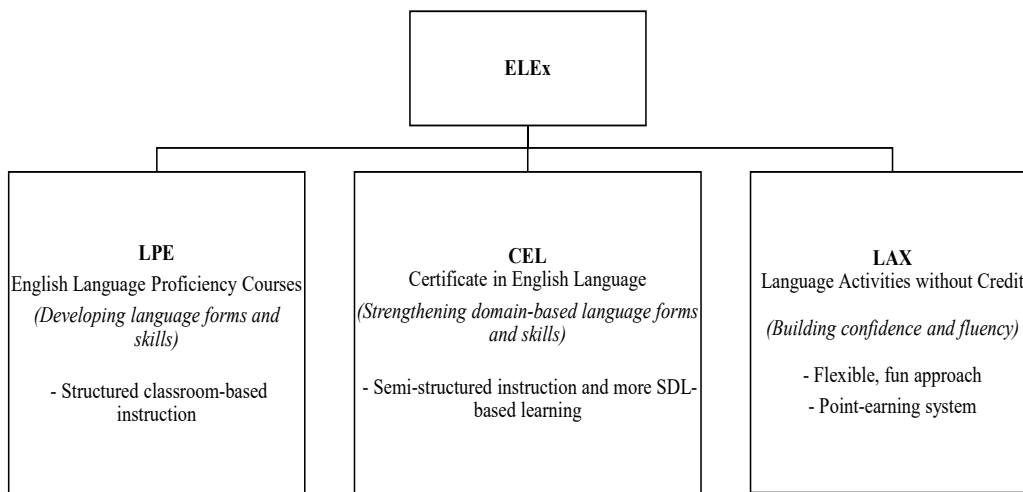

Figure 1. Overview of the ELEx Structure

The components that construct the ELEx are portrayed in Figure 1 (Centre for the Advancement of Language Competence, 2013). As displayed, the ELEx package comprises three significant parts. The LPE component focuses on building the basis of language whereby the knowledge of vocabulary and grammar would be solidified to achieve language accuracy and fluency. Similarly, the CEL component is also constructed to assist in the mastery of vocabulary and grammar. In addition, it also emphasises domain-based learning, whereby it serves to accommodate learning English for general, academic and professional purposes. On the other hand, the LAX component focuses on incidental learning via task-based activities, aiming improve students' confidence and familiarity in using the language.

Recognising that students vary in levels of proficiency from being very limited to very proficient users of English, the ELEx package is designed to cater to students' specific language needs, which is identified based on the levels that they have achieved in their MUET results (Band 1 to Band 6). MUET is a compulsory test that students have to take in order to be admitted into a university. Thus, ELEx provides students of MUET Band 1 or 2 with an intensive programme that aims to supply essential assistance to help foster their confidence in the language while assisting them to meet

their immediate needs required in academic tasks. For those with MUET bands of 3 to 6, the package would provide programmes and courses that aim to enhance further and polish their language competency as well as amplify their confidence and fluency in order to be more linguistically marketable (Abdullah et al., 2015).

The implementation of the ELEx package has proven to show positive outcomes in improving language proficiency, especially among the less proficient students, whereby the students portrayed higher willingness to use the language, which resulted in higher participation and interaction in various contexts (Mustafa, 2018). Although the ELEx package is compulsory for all students, it is seen as more of assistance for students to cope better with their studies rather than an obligation or a test (Sani, 2020). It could be said that this package delivers and is in line with the aspiration of the Ministry of higher education in developing graduates that possess adequate English language abilities. In fact, in the 14th parliament meeting on July 22, 2019, the Ministry of Education mentioned and

acknowledged the ELEx package from UPM as one of the government's efforts in assisting youths in mastering the English language (Parlimen Malaysia, 2019)

Since its implementation, the assessment for the ELEx package for each student was evaluated via an alphabetical grade for the LPE component, a 1 to 4 level for the CEL component and a Satisfactory / Unsatisfactory grade for the LAX activities. However, a comprehensive evaluation of the student's English language proficiency was not provided at the point of graduation. Therefore, the English Language Competence Score Average (ELCSA) was established, and the students will obtain scores ranging from 0.0 to 5.0. It is obtained by calculating average achievement points for two components of ELEx, namely the conventional courses (LPE) and preparatory courses (CEL). LAX activities are not included in the calculation as the activities in LAX were for scaffolding purposes that provided opportunities to use the language and build confidence. The division of courses and calculation of the student's ELCSA is shown in Table 1.

Table 1

Courses are taken into account for the determination of ELCSA according to MUET results

MUET band level	CEL courses	Number of CEL courses	LPE courses	Number of LPE courses	Total courses
1 – 2	CEL2102, CEL2103 and one of the courses CEL2105 / 2106/2107	3	LPE2301, LPE2501	2	5

Table 1 *(Continued)*

MUET band level	CEL courses	Number of CEL courses	LPE courses	Number of LPE courses	Total courses
3 – 4	CEL2103 and one of the courses CEL2102 / 2105/2106/2107	2	LPE2301, LPE2501	2	4
5 – 6	CEL2103	1	LPE2402 and / or LPE2502	1 or 2	2 or 3

The scores that will be given for the achievement of each CEL and LPE course are in Table 2 and Table 3.

Table 2

Scores for Certificate in English Language (CEL) Course

Level	Score
1	0
2	3.0
3	3.5
4	4.0

Table 3

Score for courses of Language Proficiency in English (LPE)

Alphabetical Grade	LPE2301*	LPE2501*	LPE2402**	LPE2502**
A	4.0	4.0	4.5	4.5
A-	3.7	3.7	4.0	4.0
B+	3.3	3.3	3.7	3.7
B	3.0	3.0	3.3	3.3
B-	2.7	2.7	3.0	3.0
C+	2.3	2.3	2.7	2.7
C	2.0	2.0	2.3	2.3
C-	1.7	1.7	2.0	2.0
D+	1.3	1.3	1.7	1.7
D	1.0	1.0	1.3	1.3
F	0	0	0	0

Taken only by MUET students 1 - 4

** One or both courses are taken only by MUET Students 5-6

The calculation of ELCSA is based on the total score obtained divided by the number of selected LPE and CEL courses taken (i.e., on average). The average score obtained will determine the level of ELCSA as described in Table 4. As mentioned before, LAX activities are excluded from the calculation because their main purpose is to build confidence in using English. The LPE 2401 course is also excluded because it is in special preparation for students with MUET results 1 and 2. Higher scores are given for LPE2402 and LPE2502 courses as these two courses are high-level courses taken only by MUET students 5 and 6.

Table 4

Scale for English Language Competence Score Average (ELCSA)

Score	Competency	Grade	Estimated CEFR
3.90 above	Excellent	A+	C2
3.725 - 3.89	Very High	A	C1
3.5 - 3.724	High	A-	B2
3.0 - 3.49	Competent	B+	B2
2.5 - 2.99	Average	B	B1
2.5 and below	Low	B-	B1

Linguaskill English Language Proficiency Test

Linguaskill is one of the tests provided by Cambridge Assessment English and has just recently been introduced in Malaysia in 2020. In implementing the CEFR, Cambridge Assessment English played a contributing role and possesses increasing, ongoing and various evidence that supports it to be the embodiment and reflection of the CEFR in multiple aspects (Cambridge Assessment English, 2021b). Linguaskill is a CEFR-aligned, computer-based, multi-level test that assesses one's English language proficiency in writing, reading, listening and speaking (Cambridge Assessment English, 2019). Table 5 illustrates the Linguaskill scores and corresponding CEFR levels.

Table 5

CEFR scores and levels

Cambridge English Scale Score	CEFR Level
180+	C1 or above
160–179	B2
140–159	B1
120–139	A2
100–119	A1
82–99	Below A1

The Linguaskill test provides two test options, namely Business and General, whereby Linguaskill Business assesses the familiarity of the test-taker towards the language of business. At the same time, Linguaskill General would focus on assessing English used in daily life (Cambridge Assessment English, 2019). Linguaskill Business has replaced BULATS that was officially discontinued on December 6, 2019 (Cambridge Assessment English, 2021a). While the Ministry of Higher Education Malaysia (MOHE) had specified several English competency tests (e.g., MUET, IELTS, and TOEFL iBT) that can be recognised by universities to meet English language requirements for student admission, the Linguaskill, Cambridge English Qualifications and OET was also added to the list in 2020 (Cambridge Assessment English, 2021d) Linguaskill reports up to a maximum score of 180+ which is equivalent to C1 or above on the CEFR scale (Cambridge Assessment English, 2021c). Linguaskill was developed by a team of experts and is supported by artificial intelligence. The trial report in April 2016 shows that the Linguaskill test scores are reliable and precise (Cambridge Assessment English, 2016). An analogous measure, the Rasch reliability, was used, and each test obtained a reliability coefficient over .90, which is considered adequate. Whereas the target level of precision was roughly 90% in which most of the tests that failed to reach the target precision were at the extremes of the CEFR: Level A1 or below and C1 or above (Cambridge Assessment English, 2016).

For this research, the Linguaskill General test was used. As mentioned earlier, the Linguaskill General test assesses language used in day-to-day life. The test would include topics involved with studying and working, making plans, travel and technology. Thus, it makes the test suitable for a broad spectrum of organisations, university admissions or exits. The test could also be used for recruitment roles that do not require specialist business terminology; for instance, it would be suitable for employees who are required to showcase their strong command in English to perform their roles effectively.

The Linguaskill General test has three modules which are reading and listening, speaking and writing. The reading and listening tests are adaptive according to the candidate's proficiency level, meaning that each candidate would face a different set of items on their test based on how well they answered the previous question (Cambridge Assessment English, 2018). Although there are not a fixed number of questions, each question the candidates' answer would help the computer understand their level better. The test finishes when the candidate has answered enough questions for Linguaskill to identify their level accurately. The writing test uses innovative auto-marker technology whereby the computer automatically marks it. Meanwhile, a hybrid approach was taken to mark the Linguaskill Speaking test, which uses auto-marking technology and human examiners to ensure efficiency (Xu et al., 2020).

RESEARCH METHODOLOGY

Generally, there are two approaches in which alignment to the CEFR can be adopted: the direct alignment and the indirect alignment (Bruce & Hamp-Lyons, 2015). The direct alignment would require much expertise, resources, and funding which is made possible by large organisations such as Cambridge English Assessment (Ali et al., 2018). Due to time and financial constraints, the direct approach would not be feasible. Alternatively, the indirect approach to the CEFR is adopted by mapping test scores to the CEFR-aligned scores. However,

certain factors regarding the language test such as its purpose, format, test-takers, and the scoring system should be considered before the indirect alignment can be made (Ali et al., 2018). This indirect linkage via 'equation' to an existing test already linked to the CEFR is one of the recommended approaches in the Council of Europe's Manual (Cambridge English Language Assessment, 2011).

According to the Standards for Educational and Psychological Testing (American Educational Research Association, 1999), scores can be considered 'comparable' or 'equivalent' when the test's features are closely similar to each other (Lim, 2017). In this case, this study attempts to uncover the relationship between ELSCA scores and the Linguaskill test scores. Both are designed to measure English proficiency directed towards the goal of real-world applications. Therefore, in order to fulfil the purpose of this research, a quantitative,

correlational design was utilised in this study involving the collection of quantitative data followed by a correlational analysis as the study intended to examine the extent to which two or more variables relate to one another (Rockinson-Szapkiw, 2012).

Data Collection and Analysis

The researchers were given access to the participant's ELSCA scores and their full Linguaskill test report, including their overall CEFR score and language skill scores. The researchers used Excel to compile the participant's scores accordingly and then proceeded to use the IBM SPSS Statistics software to calculate the Spearman Rho correlation to uncover the relationship between the two variables. A Spearman Rho correlation was used in this study as it can describe two variables in a monotonic relationship. It should be mentioned that the Spearman Rho correlation seems most befitting as it is suitable for data that is, either ordinal, interval and ratio variables, continuous and non-normally distributed (Schober et al., 2018). In ensuring the standard of quality when assessing the correlational analysis, outliers were addressed and removed. The presence of outliers is common in data collection due to various reasons. It, therefore, is crucial to be dealt with prior to the analysis to ensure the overall reliability of the results (Kwak & Kim, 2017). Additionally, a scatter plot was constructed to observe the relationship between the two variables further, and a trend line was identified.

Sampling

The participants were 197 final year undergraduates from six Science, Technology, Engineering and Mathematics (STEM) based faculties. Table 6 shows the participant's profile.

Table 6

Participants' profile

Participants	Category	Number of test takers	Percentage (%)
Gender	Male	41	20.8
	Female	156	79.1
Age range	21-23	153	77.7
	24 and above	44	22.3
Faculty of	Biotechnology and Biomolecular Sciences	34	17.3
	Computer Science and Information Technology	32	16.2
	Engineering	34	17.3
	Food Science and Technology	34	17.3
	Medicine and Health Sciences	30	15.2
	Science	33	16.7

As shown in Table 6, the participants were 197 final year students (M= 41, F= 156) from six STEM-based faculties in UPM. The purposive sampling method, specifically the Homogenous Sampling, was applied as this sampling form focuses on a particular characteristic of a population where they share similar traits (Etiken et al., 2016). In this case, the participants were chosen according to the following criteria; 1) Participants have completed their undergraduate programmes and therefore also obtained their ELSCA scores. 2) Participants have taken the Linguaskill test and obtained their CEFR band level. 3) Participants were among the STEM-related faculties. The number of participants from each faculty ranged from 30 to 34. The highest number of participants were from the Faculty of Engineering (n= 34), the Faculty of Food Science and Technology (n= 34), and the Faculty of Biotechnology and Biomolecular Sciences (n= 34). It is followed by the Faculty of Computer Science and Information Technology (n= 32) and the Faculty of Medicine and Health Sciences (n= 30). This study specifically chose Science Technology Engineering and Mathematics (STEM) undergraduates because based on their MUET scores, the STEM undergraduate students have varied levels of English language proficiency, which may provide better insights into the correlation between the CEFR and ELCSA

scores. For courses related to English, they are required to meet a MUET band 4 to be admitted in the course (UPM, n.d.). Furthermore, past research has shown that STEM graduates have low employment rates, possibly due to a lack of multiple skills and English proficiency (Murtaza & Saleh, 2018; Thomas, 2019). Additionally, the participants obtained both an ELSCA and a Linguaskill General score, thus allowing the comparison and correlation between ELCSA and Linguaskill.

RESULTS

The performance of the students on the ELCSA and CEFR is presented in Table 7.

Table 7

Performance of Respondents on ELCSA and CEFR

	N	Minimum	Maximum	Mean	Std. Deviation
ELCSA	197	3.000	4.000	3.667	0.211
CEFR	197	122	180	168.43	10.012

The mean ELCSA and CEFR Linguaskill scores were 3.667 and 168.430, respectively. Thus, the Linguaskill score indicates that, on average, the UPM STEM undergraduates had successfully achieved the B2 level as targeted by Malaysia's Ministry of Education.

A correlational analysis between the CEFR and ELCSA scores is presented in Table 8.

Table 8

Correlation between ELCSA and CEFR scores

			ELCSA	Overall (CEFR)
Spearman's Rho	ELCSA	Correlation Coefficient	1.000	0.371**
		Sig. (2-Tailed)	0.000	0.000
		N	197	197

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

In fulfilling research objective 2, results show a positive, weak relationship according to the Guilford Rule of Thumb between ELCSA and CEFR scores. In addition, results of Spearman Rho correlation indicated that there was a significant positive association between the overall ELCSA scores and CEFR scores, (r_s (195) = 0.371, p < .05).
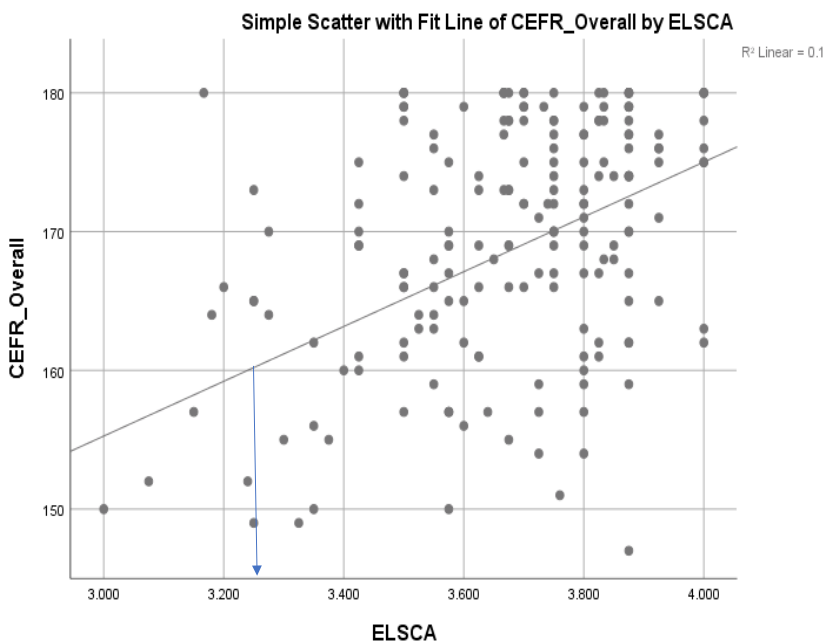
Figure 2. Simple Scatter plot of ELCSA and CEFR scores

Scores on the CEFR and ELCSA were also placed on a simple scatter plot, and based on the trend line in the scatter plot (Figure 2), a score of approximately 3.25 on the ELCSA can be considered equivalent to a Linguaskill score of 160 (CEFR Band B2). However, the trend line does not allow for predicting the C1 Band based on the ELCSA.

The correlations between the language components in the Linguaskill and the ELCSA are presented in Table 9.

Table 9

Correlation between ELCSA and CEFR scores

			Writing (CEFR)	Reading	Speaking	Listening
Spearman's Rho	ELCSA	Correlation Coefficient	0.417**	0.360**	0.249**	0.179*
		Sig. (2-Tailed)	0.000	0.000	0.001	0.012
		N	196	195	188	197
** Correlation is significant at the 0.01 Level (2-Tailed)						

Regarding the individual skills, there are correlations of varied strengths between the ELCSA and each of the four skills. There is a positive and moderate relationship between ELCSA and CEFR Writing scores with a correlation coefficient of 0.417. Also, there was a positive, low relationship between ELCSA and CEFR Reading scores with a correlation coefficient of 0.360 and CEFR Speaking scores with a correlation coefficient of 0.249. However, although positive, the relationship between ELCSA and Listening scores was negligible, with a correlation coefficient of 0.179. The relationship that was considered best and strongest was that of ELCSA and writing skills. For that reason, as well as writing being especially important in academic contexts, this relationship is further explored as in Figure 3 in order to determine the ELCSA score that would best reflect a B2 CEFR level.

The scatter plot of scores on the ELCSA, and the Linguaskill Writing skill is presented in Figure 3.
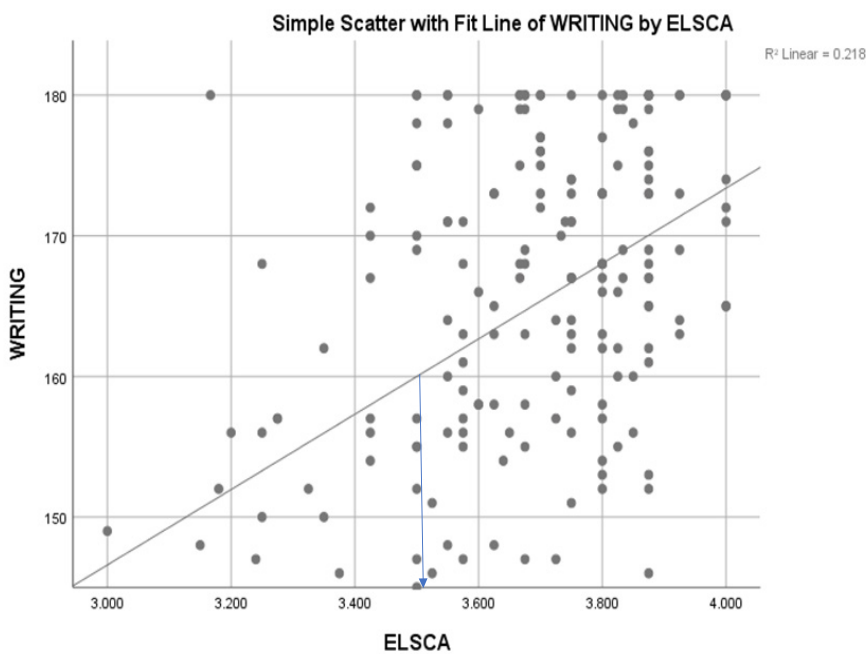

Figure 3. Simple Scatter plot of ELCSA and CEFR Writing scores

As the writing component of the Linguaskill test was the language skill that yielded the strongest correlation with the ELCSA, the scatter plot was used to identify the ELCSA score comparable to a B2 CEFR level. Based on the trend line, the ELCSA score of approximately 3.5 could be identified as equivalent to the CEFR Writing score of 160, which the Linguaskill test specifies as representing the B2 level.

DISCUSSION

The study demonstrates that there is a positive relationship between the ELSCA and CEFR scores. It means that the students that managed to get a high score in ELSCA also managed to get a high CEFR score in the Linguaskill test, and it is likewise for those who received low scores in ELSCA also possessed a low CEFR score. The correlations did, however, differ to some extent in terms of strength. The possible explanation for why the correlations varied in terms of strength is that, though the two scores both measure English proficiency for real-world applications, the two measurements' nature and grading scale differ. While the ELCSA score is cumulative based on language courses taken over time, the Linguaskill test is an English proficiency test. In terms of the grading scales, ELCSA is a score that ranges from 0.00 to 5.00, while the highest possible score obtained in the Linguaskill has a maximum score of 180, which is considered, as equivalent to a C1 and above grade on the CEFR. However, despite the varied strengths of the correlation, the data shows that a positive correlation exists, indicating that the variables move in the same direction.

Furthermore, this paper has shown that it is possible to use an established test that is CEFR aligned as a reference to determine the required scores that match a B2 level in a university English proficiency programme. For example, this study shows that a 3.25 score in the Writing component of the ELCSA corresponds to the B2 CEFR level. In comparison, a 3.5 overall score for ELCSA corresponds to the B2 CEFR level for overall English language proficiency. Thus, it could be assumed that a student who achieves a score of 3.25 in the ELCSA Writing component is at the B2 level of proficiency in terms of writing skills. Also, achieving a score of 3.5 in the ELCSA overall score would mean that a student is at B2 level for overall English language proficiency. This benchmarking is useful, as it can indicate a student's CEFR level using an internally developed university English language programme. Determining the score corresponding to the B2 CEFR level is also important as university students are expected to have a minimum B2 level of proficiency upon graduation. Notably, in so far as the students' performance is concerned, 72.6% of the STEM participants in the study managed to achieve the target that the Malaysian Ministry of Education had set by obtaining the minimum CEFR level of B2 for Malaysian university graduates. The other ten per cent of the participants had exceeded the target and managed to achieve C1, while only 17.3% achieved B1 and fell below the Ministry target.

Previous studies had mentioned that caution should be taken when aligning assessments using CEFR as it was implied that although the different tests use related criteria and are based on descriptors of the same however the perceived equivalence is only assumed (Foley, 2019). Additionally, it should also be considered that even though tests such as IELTS has been aligned to the CEFR, the alignment does not refer to the scores of specific language skill; instead,

it refers to the overall scores (Ali et al., 2018). Nonetheless, a study conducted by Universiti Malaysia Pahang attempted to contextualise the CEFR with their English Writing Language Proficiency Test. Their preliminary analysis has shown that the CEFR-A1 is sufficient in describing their lowest band (Band 1) and that the CEFR C2 and C1 would describe their highest bands, namely band 8 and 9. It was also mentioned that it was necessary to further describe the subcategories of the level of proficiency in order to address all of their bands as their English proficiency test had nine bands altogether (Ali et al., 2018). Therefore, it could be said that, despite being cautious of comparability aspects and over emphasis on standardisation, attempts for an alignment can be made possible. However, it is important to note that fundamentally, the CEFR was originally devised to assist the planning of curricula and that the common reference levels are for further facilitation (Foley, 2019).

CONCLUSION

In seeking to align the accumulative ELSCA scores with the Linguaskill CEFR scores, the authors conclude that there is a positive correlation between the ELSCA scores and the CEFR scores—which shows that there is a possibility in using performance in an English language proficiency programme to predict CEFR levels. Furthermore, this study has also shown that the ELCSA can be used with either the Linguaskill overall score or the writing score to predict and determine CEFR levels, especially to

indicate whether or not the student has attained B2 in the CEFR as required by the Ministry of Education for university students upon graduation. Due to this alignment, it can be said that UPM is on the right track in benchmarking its language proficiency programmes with the CEFR. However, it is important to ensure the efficiency of their language programmes and make improvements where necessary.

It is suggested that for future research, attempts should be made to benchmark language programmes in different higher learning institutions to the CEFR. Given that the Linguaskill test is now accepted and adopted in the admission and exit requirements of universities in Malaysia as an alternative to MUET, IELTS, TOEFL and other tests, language centres should consider providing training for students to prepare for such tests or even become centres to carry out the tests. It could further enhance the curriculum of language centres and the practices of language instructors to be more CEFR-aligned. Consequently, this would increase the student's familiarisation with the CEFR and help them develop their language proficiency in line with the CEFR. In sum, this paper contributes knowledge that an alignment between a language proficiency programme of a Malaysian university and the CEFR does exist and that it is pertinent for other institutes to work in unanimity to benchmark their language proficiency programs towards the CEFR so that the level of standards of the English Language in Malaysian universities are acceptable and further credible.

ACKNOWLEDGEMENT

REFERENCES

Abdullah, M. H., Wah, H. S., & Eng, W. B. (2015). Delivering the ELEx package: An English language experience approach for developing undergraduates' language proficiency. *Journal of Language and Communication, 2*(2), 283-296.

Ali, Z. M., Ali, F., Radzuan, N. M., Alwi, N. N. M., Abu, N. L., & Kassim, Z. (2018). Contextualising the CEFR: the Universiti Malaysia Pahang English language proficiency writing test. In *11th Annual International Conference of Education, Research and Innovation (ICERI 2018)* (pp. 4892-4902). IATED Publisher.

Alosani, M. S., Al-Dhaafri, H. S., & Yusoff, R. Z. (2016). Mechanism of benchmarking and its impact on organizational performance. *International Journal of Business and Management*, *11*(10), 172-183. https://doi.org/10.5539/ijbm.v11n10p172

American Educational Research Association. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Bogan, C. E., & English, M. J. (1994). *Benchmarking for best practices: Winning through innovative adaptation.* Mcgraw-Hill.

Bruce, E., & Hamp-Lyons, L. (2015). Opposing tensions of local and international standards for EAP writing programmes: Who are we assessing for? *Journal of English for Academic Purposes, 18*, 65-77. https://doi.org/10.1016/j.jeap.2015.03.003

Cambridge Assessment English. (2016). *Linguaskill.* Cambridge Assessment English. https:// www.cambridgeenglish.org/Images/429191-linguaskill-listening-and-reading-trial-report.pdf

Cambridge Assessment English. (2018). *Meet the research team behind Linguaskill.* Cambridge Assessment English. https:// www.cambridgeenglish.org/Images/521541-linguaskill-research-team.pdf

Cambridge Assessment English. (2019). *Linguaskill simplify your English language testing.* Cambridge Assessment English. https:// www.cambridgeenglish.org/Images/466596-linguaskill-simplify-your-english-language-testing.pdf

Cambridge Assessment English. (2021a). *BULATS*. Cambridge Assessment English. https://www.cambridgeenglish.org/exams-and-tests/bulats/

Cambridge Assessment English. (2021b). *Fitness for purpose*. Cambridge Assessment English. https://www.cambridgeenglish.org/research-and-validation/fitness-for-purpose/#cefr-relationship

Cambridge Assessment English. (2021c). *How results are presented*. Cambridge Assessment English. https://www.cambridgeenglish.org/exams-and-tests/linguaskill/information-about-the-test/how-results-are-presented/

Cambridge Assessment English. (2021d). *Malaysia: Ministry of Higher Education recognises English language exams from Cambridge*. Cambridge Assessment English. https://www.cambridgeenglish.org/news/view/malaysia-ministry-of-higher-education-recognises-cambridge-english-exams/

Cambridge English Language Assessment. (2011). *Using the CEFR: Principles of Good Practice.* University of Cambridge. https://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf

Canadian Language Benchmarks. (2012). *English as a second language for adults.* Centre for Canadian Language Benchmarks. https://www.canada.ca/

content/dam/ircc/migration/ircc/english/pdf/pub/language-benchmarks.pdf

Centre for the Advancement of Language Competence. (2013). *ELEx Package (English language experience) for undergraduates* Centre for the Advancement of Language Competence. https://calc.upm.edu.my/dokumen/11203_ELEX_Briefing_BI.pdf

Cheung, H. (2012). Academic perspectives from Taiwan. In M. Byram, & L. Parmenter (Eds.), *The Common European Framework of Reference: The globalisation of language education policy* (pp. 224-230). Multilingual Matters.

Don, Z. M., & Abdullah, M. H. (2019, May 27). What the CEFR is and isn't. *Free Malaysia Today*. https://www.moe.gov.my/en/menumedia/printed-media/newspaper-clippings/what-the-cefr-is-and-isn-t-free-malaysia-today-27-mei-2019

Etiken, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

Fisher, J. G. (1996). *A managers guide to staff incentives and performance improvements.* Kogan Page Publication.

Foley, J. A. (2019). Issues on assessment using CEFR in the region. *Language Education and Acquisition Research Network Journal, 12*(2), 28-48.

Herman, J. L., & Osmundson, D. R. (2010). *Benchmark assessment for improved learning (AACC Policy Brief).* University of California. https://doi.org/10.1037/e685212011-001

Inguva, M. (2018). Effective use of benchmarking: The context of the centre for preparatory studies in Oman. *Arab World English Journal*, 93-104. https://doi.org/10.24093/awej/MEC1.7

Jaafar, M. S. (2019). *Syarat kompetensi Bahasa Inggeris kepada pelajar antarabangsa* [English language competency requirements for international students]. Malaysian Qualifications Agency. https://www.mqa.gov.my/pv4/document/pekeliling/2019/JPT-Penambahbaikan%20Syarat%20Kompetensi%20BI%20Pelajar%20Antarabangsa%20IPTS%2009Dis.pdf

Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the common Chinese Framework of Reference for Languages: Challenges at macro and micropolitical levels. *Language Testing in Asia*, *7*(1), 1-19. https://doi.org/10.1186/s40468-017-0032-5

Keegan, R., & O'Kelly, E. (2012). *Applied benchmarking for competitiveness: A guide for SME owner/ managers.* Oaktree Press.

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: Management of missing values and outliers. *Korean Journal of Anesthesiology*, *70*(4), 407-411. https://doi.org/10.4097/kjae.2017.70.4.407

Lim, G. S. (2017). *The Occupational English Test and IELTS: A Benchmarking Report.* Cambridge Assessment English.

Malaysian Qualifications Agency. (2020). *Kompilasi dasar jaminan kuasa pendidikan tinggi (2009-2020): Edisi* ketiga [Compilation of higher education power guarantee policies (2009-2020): 3rd ed.]. https://www.mqa.gov.my/pv4/document/compilationPolicy/KOMPILASI%20DASAR%20JAMINAN%20KUALITI%20PENDIDIKAN%20TINGGI%2020092020.pdf

Malaysian Qualifications Agency. (2020). *Malaysian qualification agency.* Programme Standards: Language. https://www2.mqa.gov.my/qad/v2/garispanduan/2020/PS%20Language%20-%20updates%20Oct%202020.pdf

Markel, S. (2018). The CEFR and English education in Japan. *Journal of Policy Studies*, *56*, 33-38.

Ministry of Education. (2015). *English language education reform in Malaysia: The Roadmap 2015-2025.* Ministry of Education Malaysia.

Murtaza, S. F., & Saleh, N. S. (2018). English language use in Malaysian government and private civil engineering workplaces. *International Journal of Education & Literacy Studies*, *6*(3), 84-91. https://doi.org/10.7575/aiac.ijels.v.6n.3p.84

Mustafa, Z. (2018, July 11). Proficient way to better English. *New Strait Times*. https://www.nst.com.my/education/2018/07/389562/proficient-way-better-english

New Strait Times. (2019, June 27). *We can make CEFR work for us*. *New Straits Times*. https://www.nst.com.my/opinion/letters/2019/06/499490/we-can-make-cefr-work-us

Nwabuko, L. O., Igwe, N. J., Okengwu, M. C., Nwabuko, M. A., & Ekere, O. (2020). Benchmark for partnership in human and material resources provision for adult education programmes in the South-East zone of Nigeria. *Global Journal of Health Science*, *12*(5), 46-56. https://doi.org/10.5539/gjhs.v12n5p46

Parlimen Malaysia. (2019). *Mesyuarat Kedua, Penggal Kedua Parlimen Keempat Belas 2019* [Second Meeting in the Second Term of the 14th Parliamentary]. Parlimen Malaysia. https://www.parlimen.gov.my/files/jindex/pdf/JDN22072019.pdf

Prakash, G. (2019, May 23). Education Ministry says using European benchmark for English-language option teachers. *Malay Mail*. https://www.malaymail.com/news/malaysia/2019/05/23/education-ministry-says-using-european-benchmark-for-english-language-optio/1755690

Read, J. (2019). The influence of the Common European framework of Reference (CEFR) in the Asia Pacific region. *LEARN Journal: Language Education and Acquisition Research Network Journal*, *12*(1), 12-18.

Rethinasamy, S., & Chuah, K. M. (2011). The Malaysia university English test and its use for placement purposes: A predictive validity study. *Electric Journal of Foreign Language Teaching*, *8*(2), 234-245. https://doi.org/10.2139/ssrn.2146007

Rockinson-Szapkiw, A. J. (2012). *Step 4: Selecting and justifying your research design.* http://amandaszapkiw.com/artifacts/resources/tutorials/research-process/Step-4-Selecting-and-Justifying-Your-Research-Design.pdf

Sani, R. (2018, February 7). *Building capacity to increase English Proficiency*. *New Strait Times*. https://www.nst.com.my/education/2018/02/332968/building-capacity-increase-english-

Sani, R. (2020, March 18). *Meeting university requirements*. *New Strait Times*. https://www.nst.com.my/education/2020/03/575837/meeting-university-requirements

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Sufi, M. K. A., & Stapa, M. (2020). Should the CEFR illustrative scales be localised to Malaysian higher education standards? – A conceptual paper. *International Journal of Academic Research in Business and Social Sciences, 10*(9), 885-897. https://doi.org/10.6007/IJARBSS/v10-i9/7882

Thomas, J. (2019, November 28). *Malaysia: Between education and skills*. *The Asean Post*. https://theaseanpost.com/article/malaysia-between-education-and-skills

UPM. (n.d.). *Specific Program Requirements for Bachelor Program*. Universiti Putra Malaysia.

https://fsktm.upm.edu.my/upload/dokumen/BGAKA1_Specific_Program_Requirements_(Bachelor).pdf

Wu, J. (2012). Policy perspectives from Taiwan. In M. Byram, & L. Parmenter (Eds.), *The Common European Framework of Reference: The globalisation of language education policy* (pp. 213-223). Multilingual Matters.

Xu, J., Brenchley, M., Jones, E., Pinnington, A., Benjamin, T., Knill, K., Seal-Coon, G., Robinson, M., & Geranpayeh, A. (2020). *Linguaskill Building a validity argument for the Speaking test.* Cambridge Assessment English. https://www.cambridgeenglish.org/Images/589637-linguaskill-building-a-validity-argument-for-the-speaking-test.pdf

Zairi, M. (1996). *Benchmarking for best practice: Continuous learning through sustainable.* Heinemann.

Zulkefli, A. S. (2017, November 4). M'sian students, teachers show vast improvement in CEFR as per Cambridge Evaluation Study 2017. *New Straits Times*. https://www.nst.com.my/news/nation/2017/11/299126/msian-students-teachers-show-vast-improvement-cefr-cambridge-evaluation