Developing Wordlist for Primary Science Textbooks in Dual Language Programme

Nur Ilyani Idris[a,b], Nor Liza Ali[b], Fazilah Jaafar[a]

[a]Faculty of Education, Humanities & Arts, Universiti Poly-Tech Malaysia, Kuala Lumpur, Malaysia, [b]Faculty of Social Sciences & Humanities, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

Abstract
Dual Language Programme (DLP) is an initiative programme to produce students who can understand and communicate in English language. The implementation of the programme in a school is optional, which means the decision to have a DLP in a particular school is left to the school headmaster and parents, as parent groups played a critical role to bring back English as a medium of instruction in the national school system. However, one of main issues in the DLP implementation is students' linguistic readiness to learn in English, particularly students in primary school. This paper discusses how language used in science textbooks (Years 1-6) were analysed using the readability reference manager Text Inspector to generate the wordlist. Thus, this study examines the readability of the textbooks used in primary school to understand the level of difficulty of the language. The preliminary finding showed that the readability of primary science textbooks is way beyond CEFR A2. Hence, this study proposes to generate a wordlist for science textbooks in DLP in primary schools based on CEFR readability levels.
Keywords: Dual Language Programme, Science, Readability, Wordlist, CEFR

Introduction
In Malaysia, the interest in the use of English as a medium of instruction (EMI) at school level has continued to meet with the modernisation and globalisation especially in the application of Science, Technology, Engineering and Mathematics (STEM). Hence, the Ministry of Education (MoE) has implemented a new policy *Memartabatkan Bahasa Malaysia Memperkukuh Bahasa Inggeris* (MBMMBI) in 2009 (Ministry of Education Malaysia, 2014). In the policy, Dual Language Programme (DLP) was introduced, where English is used as a medium of instruction in the teaching of science and mathematics at primary and secondary schools. The aim of this programme is to provide more exposure to English language; thus, will help to improve the level of English proficiency among students (Ministry of Education Malaysia, 2015). DLP is an optional programme which the decision to implement the programme or not is left to school headmasters to decide. In addition, parents also play an

important role to decide the medium of instruction for the children (Ministry of Education Malaysia, 2015).

To implement the programme, schools need to fulfill the criteria stated by the MoE: certification by headmasters that the teachers are able to teach science and mathematics in English; adequate classrooms; written requests from parents for DLP classes; satisfactory achievement in Bahasa Malaysia in the national exam or achieving above the standard level specified by the MoE; and at least 15 students per class. In addition, to qualify to teach in English, teachers are required to obtain at least C in English in Sijil Pelajaran Malaysia (SPM) or its equivalent (Ministry of Education Malaysia, 2020).

Even though this policy provides the flexibility to schools to choose the medium of instruction to learn science and mathematics subject in English, there are issues in the implementation such as on teachers' linguistic readiness as well as students' linguistic readiness. Language is a carrier of knowledge and learning. Having inadequate proficiency in English may hinder learning especially in subjects such as mathematics and science. Besides, learning science itself is not easy and very challenging especially to understand the concepts in English (Ali & Ismail, 2006; Masrom, Idris & Jusoh, 2021). Students with limited English proficiency will face difficulties in comprehending the content of the subject in which the words or terminologies are not readable to them. Although limited, there are studies that examined readability of science textbooks in their contexts (see e.g., Ayodele, 2013; Francis et al, 2020)

This study conducted a preliminary analysis of a selected topic from the science textbooks from Year 1 to Year 6, and the findings indicate that the readability levels of the science texts are beyond CEFR A2 level. The authors acknowledge that this level of difficulty is word level, not exactly the content level as the information in the textbooks are presented with graphical images to help comprehension. However, the authors feel that there is a need to create the awareness on the word difficulty particularly to science teachers, so they can decide which words require explanations to help comprehension.

Hence, this study proposes to generate a wordlist for science textbooks used in DLP for primary schools in Malaysia based on Common European Framework of Reference for Languages (CEFR) readability levels. Selected topics from science textbooks for each grade level (Years 1-6) were selected, processed and analysed using the Text Inspector webtool. The use of CEFR is also in view of the introduction CEFR in English language education in Malaysia in 2017.

Common European Framework of Reference (CEFR)
The Common European Framework of Reference (CEFR) was introduced as part of an initiative to improve English language education in Malaysia. The adoption of the CEFR is an aspiration for the prevailing international standard of English proficiency of the students from pre-school to post-secondary education, as well as English proficiency of English teachers. The aspiration is substantive and the plan for the implementation is documented in the English Language Education Reform in Malaysia: The Roadmap 2015-2025 (Ministry of Education Malaysia, 2015).

The CEFR is a framework of standard to reflect the level language proficiency of any language, not just English. Nevertheless, scholars have used the descriptive standard in the CEFR to many aspects of language education, mainly curriculum, assessment, methodology (Council of Europe, 2018). In addition, CEFR is also used to determine text difficulty such as Flesch Kincaid, Text Inspector and others. Traditionally, language learners have always

categorised as beginners, intermediate, and advanced learners. However, CEFR identified learners into three levels: *Basic* (A1 and A2), *Independent* (B1 and B2), and *Proficient* (C1 and C2). This framework provides positive descriptors to asses learner's ability with "can do" statements rather than what the students are unable to do (Ministry of Education Malaysia, 2015; Mohd Don & Abdullah, 2019). Table 1 shows the CEFR descriptors, which indicate the ability that a learner can use the language according to the CEFR levels.

Table 1
Overall Reading Comprehension (Council of Europe, 2018; p.60)

Overall Reading Comprehension	
C2	Can understand virtually all forms of the written language including abstract, structurally complex, or highly colloquial literary and non-literary writings Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.
C1	Can understand in detail lengthy, complex texts, whether or not they relate to his/her own area of speciality, provided he/she can reread difficult sections. Can understand a wide variety of texts including literary writings, newspaper or magazine articles, and specialized academic or professional publications, provided that there are opportunities for re-reading and he/she has access to reference tools.
B2	Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.
B1	Can read straightforward factual texts on subjects related to his/her field and interests with a satisfactory level of comprehension.
A2	Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency every day or job-related language. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared
A1	Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.
Pre-A1	Can recognise familiar words accompanied by pictures, such as a fast-food restaurant menu illustrated with photos or a picture book using familiar vocabulary.

The roadmap 2015-2025 has set aspirational targets for Malaysian students to be achieved by the year 2025. This is to track student's English progress from pre-school to university (see Figure 1).

Table 2
CEFR Aspirational Targets

Education Level	CEFR Level
Pre-school	A1
Primary school	A2
Secondary school	B1
Post-secondary school	B2
University	B2/C1

Figure 1 illustrates the aspirational target that has been set by Ministry of Education Malaysia (MoE) (Ministry of Education Malaysia, 2015). The target for primary school students is CEFR A2. By the end of the primary level (Year 6), students should be able to understand short, simple texts on familiar matters of a concrete type which consist of high frequency every day or job-related language and also to understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared. Therefore, these minimum expectations can provide as a guideline to evaluate student's progress in English language.

Given that the target level for English at the end of primary school is CEFR A2, and our preliminary study on language used in the science textbooks in DLP at primary school showed that the language used is of CEFR B2 and C1 level, this raises the need to support DLP teachers and students in terms of linguistic in managing teaching and learning in EMI programme. The CEFR B2 and C1 is the aspiration target level for students upon graduating from the university. The preliminary study indicates that the students, even in Year 1, have to learn science in English of the language level of post-secondary schools and university, which this may affect their comprehension and interest in learning science.

Literature Review
Overview on the Readability Formulas
Readability is typically concerned with the suitability of a text for intended readers at a particular reading level, whether the text is too difficult or too easy to read (Zamanian & Heydari, 2012). Readability formulas measure how easy it is to read and understand a written text, based on several aspects of a text such as number of syllables in a word, number of words in a sentence, number of sentences, and whether sentences are simple or complex (Text Inspector, n.d). Among the most widely used readability formulas are Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog, and Text Inspector.

Flesch Kincaid Reading Ease is measured according to a ratio of total words, sentences and syllables. This formula gives a text a score between 1 and 100. The highest score, the easier the text, while difficult text will score lower (below 40). Flesh Reading Ease score is based on two factors: sentence length and word length. Flesch-Kincaid Grade Level readability formula provides the approximation reading grade level (equivalent to United States (US) education level) of a text, in other words, the score indicates the required education level to understand a given text (Readable, 2011).

Another formula is Flesch Kincaid Reading Grade. This formula is measured based on the average number of syllables per word and words per sentence. This readability formula is

used to assess the approximate reading grade level of a text, and it is equivalent to the United States grade level of education. For example, if a text score at 12, it means that 12th grade students (18 years old) will understand the text.

In addition, Gunning Fog is a formula to measure the readability of a text as well as text difficulty. Gunning Fog readability score is based on a formula that estimates the education level required to understand the text. The formula includes the number of words, sentences, simple sentences and complex sentences in the analysis. The analysis can be done without software (Readable, 2011). The Gunning Fog index starts from 6 to 17 and it estimates the education grade (equivalent to US education level) for the reader to understand the text. An index of 8 is readable for eight-graders, while an index above 12 is suitable for college readers.

Even though Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, and Gunning Fog are popular readability scores used by teachers to measure the readability to design tests, many not realise that those readability formulas do not follow the school grade level in English as Second Language (ESL) setting, but it follows the school grade in the United States (US) and not suitable to be used in ESL setting especially Malaysia. In addition, those formulas also do not measure the readability of the text according to CEFR level. Hence, this study used Text Inspector to measure the readability of the text in accordance to CEFR level used in primary school in Malaysia.

Text Inspector

Text inspector is a linguistic analysis tool to measure the difficulty level of English language text according to the Common European Framework of Reference (CEFR). It has several features such as Statistics, Lexical Diversity, Lexis: EVP (English Vocabulary Profile), Lexis: BNC (British National Corpus), Lexis: COCA (Corpus of Contemporary American English), Lexis: AWL (Academic Word List & Phrases), Scorecard and others to examine the readability of a text (Text Inspector, n.d).

In Text Inspector, the analysis generated by the Scorecard and Lexis: EVP is based on established metrics of readability. Each features generate statistics or analysis based on the individual needs to get comprehensive information regarding complexity, readability, lexical diversity, estimated CEFR level and other key statistics from any given text. Thus, the usefulness of Text Inspector in generating the readability score in accordance to the CEFR level has given an insight to analyse the word difficulty used in science textbook in Malaysian education setting.

Related Studies

There is a limited amount of literature on readability, but these studies show that these and other readability formulas are useful. The studies highlighted the language difficulty of the teaching materials in EMI programme and the concerns on teachers and students' linguistic readiness. Studies have shown that students have difficulties learning science in English due to the use of words or vocabularies in textbooks that are not understandable to the students.

A study conducted by Francis et al (2020) examined the readability levels of the *New Integrated Science for Junior High Schools (3rd Edition)-Book 1-3* in Junior High Schools in Ghana. In the study, two readability formulas: i) Flesch Readability Ease and ii) Flesch-Kincaid Readability Formulas were used to examine the readability level of the textbooks. In addition, cloze test was administered to 135 students to examine their comprehension level. It is found that the science textbooks used are too difficult for the students' respective levels. Even though the result showed that Textbook 3 meet the students' comprehension level, the other

2 textbooks were written above the students' level and guidance from teachers are needed to assist students in comprehending the texts.

Similarly in Ghana, Gyasi (2013) conducted a study to analyse the readability of science textbooks for senior high school from five schools. In this study, Gunning Fog Scale was used to assess the readability level of four science textbooks (Physics, Chemistry, Biology, and Integrated Science). To measure students' reading ability, Cloze Test Readability Index was used and a total of 300 high schools' students were selected to answer the test. The findings revealed that, only Chemistry textbook was appropriate for senior high school students, while Physics, Biology, and Integrated Science textbooks were too demanding for them. Based on the cloze test result, majority students experienced frustration in reading Physics, Biology, and Integrated Science textbooks, and not Chemistry textbook. Therefore, the finding emphasised that students could not understand certain terms in those textbooks and need teacher's assistance.

Ayodele (2013) examined the readability of Biology textbook and comprehension level of students in Senior Secondary School Biology in Ekiti State, Nigeria. To measure the readability level of the textbooks (*Essential Biology for Senior Secondary Schools Books 1-3*), Flesch Reading Ease Formula was used. In addition, a total of 108 of Senior Secondary School 1-3 levels were selected to answer the Cloze Test. It is found that the Biology textbooks for Senior Secondary 1 to 3 are too difficult for the intended students' level. Furthermore, cloze test results also indicated that out of 108 students, 50 students read the textbooks with frustration level, where the Biology textbooks were written above the level of comprehension expected by the students. This finding highlighted that the reading materials should be written in accordance to students' readability level to ensure the usefulness of the texts.

The review emphasised the issue of language difficulty in teaching materials, specifically textbooks, and highlighted the problem of comprehension due to the use of vocabulary and phrases beyond the level of English proficiency of students. Furthermore, these studies demonstrated the usefulness of the readability formulas (Flesch-Kincaid Reading Ease, Flesch-Kincaid Grade Level, and Gunning Fog) in explaining why some texts are easier to read than others, as well as whether the texts are appropriate for different levels of education.

It is important to note that the readability scores of the formulas are determined by the educational level in the United States. Hence, when assessing the readability of a text outside of the United States (US), one must be aware that Grade Level is correlated with the Grade Level (in terms of English) of the context, for example, pre-tertiary students from the US and Malaysia do not have the same level of language proficiency. In addition, none of these readability formulas can identify word-level difficulty, so the Text Inspector readability formula was selected for this study.

Methodology

This study examines the readability level of primary science textbooks from Year 1 to Year 6 in DLP. This study is part of a larger project. For this paper, selected units/topics of science textbooks from each grade level (Year) were analysed. Readability reference manager Text Inspector (with subscription) is used to examine the readability of the selected topics.

This study uses two features of the Text Inspector which are i) Scorecard and ii) Lexis: EVP. The Scorecard generates the lexical profile score in CEFR (e.g., CEFR C2) for the readability of the text. The lexical profile of the Scorecard is based on multiple metrics: statistics/syllables, lexical diversity, English Vocabulary Profile (EVP), British National Corpus

(BNC), Corpus of Contemporary American English (COCA), Academic Word List (AWL), and metadiscourse markers.

Lexis: EVP analyses the text according to the English Vocabulary Profile (EVP) developed by Cambridge University Press. The EVP is a reference that contains information about which words, phrases, idioms and collocations are used at each level of English learning. Lexis: EVP identifies each word used in the text according to the CEFR on a scale of A1-C2 (see Figure 2).

The analysis of Text Inspector is not reliable for short documents (below 100 words). However, primary school science textbooks do not contain many long paragraphs. The layout of the text contains graphics and short description, which the text is not in paragraph form of more than 100 words (see Figure 1). To meet the requirement for the analysis, the researcher put together the sentences of the selected to make the 100-words paragraph to be analysed in Text Inspector.
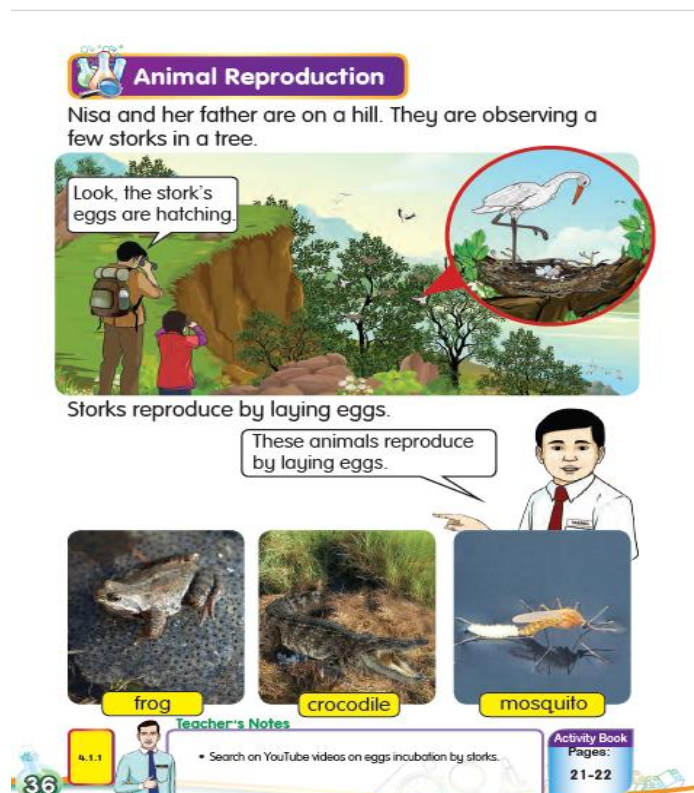
Figure 1: *An example of typical layout in science textbook (Year 2 – Unit 4; Topic – Animals)*

Figure 1 shows the content in Year 2 science textbook contains graphics and short sentences. In fact, this layout presentation is common in other grade level science textbooks.

Figure 2 shows an example on the application of the Text Inspector: Lexis: EVP on input text. The input sample is taken from Year 2 - Unit 5. Each word is tagged with the CEFR level according to EVP (see Figure 2). The percentage of words according to CEFR levels can also be generated.

Figure 2: Input Tagged Word according to CEFR Level

The above shows analysis generated by Lexis: EVP. The analysis identifies each word in accordance to the CEFR level. For the purpose of presentation for this paper, only one unit/topic for each grade-year (Years 1 to 6) was analysed.

Findings and Discussion
Readability of Science Textbooks
The preliminary study shows that the language used in the science textbooks in DLP is too demanding for primary students. Table 3 shows the CEFR level of reading difficulty according to grade (Year) level based on Text Inspector: Scorecard.

Table 3
Reading difficulty according to CEFR level

Year	Unit	CEFR level
Year 1	Unit 5-Animals	B2+
Year 2	Unit 4-Animals	B1+
Year 3	Unit 4-Animals	B2
Year 4	Unit 3-Animals	C1
Year 5	Unit 3-Animals	B2+
Year 6	Unit 3-Microorganism	C1

The selected unit for each grade level (Year) was analysed using Text Inspector: Scorecard. The results of the study indicated that all primary science texts, even for primary Year 1, have a readability level above CEFR A2. These findings suggest the difficulty of students to learn the subject, which the subject is commonly assumed difficult, but also the language used is way beyond primary school students' proficiency level. According to CEFR aspirational

targets, the target of the English language programme for primary education is CEFR A2 (see Table 2). However, this finding showed that the level of English required to learn science at primary school is beyond CEFR A2, and this leads to the concern on the selection of the textbooks, particularly on the linguistic aspects in the presentation of contents. In addition, science teachers need to be aware of this language aspect during teaching.

Generated wordlist from science textbooks

Language used in science textbooks in DLP in primary schools from Years 1 to 6 was analysed using the Text Inspector. The data presented the list of words and its percentage generated from the Text Inspector – Lexis: EVP based on selected units. This section discusses the analysis in the following order: i) CEFR A1 and A2 words; ii) CEFR B1 words; iii) CEFR B2 words; iv) CEFR C1 words; v) CEFR C2 words; and vi) *Unlisted* words.

CEFR A1 and A2 words

Table 4 shows the wordlist of A1 and A2 words generated by the Text Inspector from Years 1 to 6.

Table 4

List of words and percentages of CEFR A1 and A2

Year/ CEFR	A1 words	Percentage	A2 words	Percentage
Year 1	about, above, an, and, animal, animals, are, bird, body, can, cat, cow, different, every, feet, find, fine, fish, fly, for, from, hair, hard, has, have, head, help, horse, important, in, is, its, learn, of, other, part, parts, pictures, swimming, that, the, there, to, we, with, you	53.49	collects, duck, explain, let us, magazines, rabbit, several, themselves	9.30
Year 2	a, about, above, adult, after, also, am, an, and, animal, animals, are, are called, as, baby, be, being, birds, can, cat, cats, change, changes, come from, cow, different, do, does, eaten, egg, eggs, father, fishes from, have, her, hi, his, how, I, in, into, is, is called, it, its, life, look, make, many, milk, my,	49.13	bear, become, by, chicken, elephant, elephants, few, finally, following, going to, grow, hill, lay, let, look after, rabbits, so, such, wow	10.98

name, not, number, of, on, or, other, parents, pictures, places, safe, say, see, some, sometimes, that, the, their, them, there, these, they, this, to, tree, we, well, what, where, will, with, you, young

Year 3	about, above, all, always, and, animals, any, are, as, banana, be, because, below, big, both, but, can, can't, cannot, changed, changes, choose, cold, did, different, difficult, do, easier, eat, eaten, eating, examples, find, fish, food, for, give, grass, had, have, how, I, in, is, it, know, leaves, live, me, meat, my, name, not, of, on, only, other, pictures, plants, same, saw, teacher, teeth, than, that, the, their, there, these, they, think, this, to, we, what, which, why, would, yes, you	54.42	bears, by, carrots, chicken chickens, could, covered, cut, easily, elephant, explain, grow, however, ice, if, lion, might, monkey, rabbit, rabbits, snake, strong, suit, types	16.33
Year 4	a, about, above, all, also, am, an, and, animals, are, as, be, because, below, between, bird, birds, body, both, but, can, cat, cold, conversation, country, cows, do, evening, example, examples, fine, fish, fly, for, found, hair, hard, has, have, having, hi, horse, I, in, is, it, its, know, live, lives,	46.53	around, bats, chicken, covering, difference, explain, following, insects, lucky, most, order, snake, spend, such, through, way, were, while, wow	9.41

	more, mr, my, need to, no, not, of, often, oh, on, one, only, or, other, our, picture, same, sea, some, swim, than, that, the, their, there, these, they, this, those, time, to two, us, use, using, very, water, we, what, when, which, why, with, you			
Year 5	a, above, after, also, an, and, animal, animals, are, as, bad, be, because, below, big, black, bodies, body, can, catch, come, did, do, easy, eyes, fast, feel, fish, for, found, from, gone, hard, have, horses, how, I, is, it, leg, me, mr, my, near, not, of, often, on, or, other, our, parts really, same, some, that, the, their, them, there, these, they, this, those, three, times, to, up, used, we, what, when, why, will, with, you	38.38	able, actually, become, by, elephants, however, hurt, if, itself, kill, let us, move, out, quickly, rabbits, save, size, so, strong, such, themselves, trying, while	11.62
Year 6	a, above, all, also, an, and, animal, animals, are, as, at, bad, be, because, bodies, body, bread, can, cannot, cheese, cold, did, different, do, does, doesn't, each, easier, examples, eyes, father, find, food, for, found, from, group, groups, has, have, help, how, I, in, inside, is, it, know, known, last, leaves, live, living, longer, lot, make,	36.79	against, air, area, around, by, called, circle, dried, easily, explain, form, fresh, grow, if, kept, lakes, latest, low, matter, mean, meaning, means, medicine, mixed, most, move, neck, own, soft, still, such, temperature, through, try, turn, type, types, useful, yogurt	13.04

met, milk, mouth, mr, needed, not, of, on, one, only, or, other, our, plants, rice, rivers, seen, small, smallest, some, sugar, supermarket, than, that, the, their, them, these, they, thing, things, thinking, this, to, too, tooth, us, used, uses, using, very, warm, was, water, well, what, when, which, why, will, with, word, words, you

Table 4 showed the list of words according to the CEFR level that generated by Text Inspector-Lexis: EVP. The findings showed that the 100-words paragraph for each grade-levels (Years 1 to 6) that the highest percentage of CEFR level A1 words can be found in all Year 1 to Year 6 textbooks. This is predictable as the words at level A1 (e.g., animal, difficult, eating, important, mouth) are common and familiar in sentences. The percentages of CEFR A1 words for Year 1 (53.49 per cent), Year 2 (49.13 per cent), Year 3 (54.42 per cent), Year 4 (46.53 per cent), Year 5 (38.38 per cent) and Year 6 (36.79 per cent). The findings indicate that more CEFR A1 words are used in textbooks in lower primary (Year 1 to Year 3); however, this analysis is not conclusive as the sample for this analysis is considered small. Nevertheless, this finding adds an aspect for further investigation.

One may assume that the percentage of CEFR A2 words used in a text for each level may be slightly lower than of CEFR A1. However, the findings indicate that the percentages for CEFR A2 words used at all grade levels are significantly lower than of CEFR A1 words. The percentages of CEFR A2 words for Year 1 (9.30 per cent), Year 2 (10.98 per cent), Year 3 (16.33 per cent), Year 4 (9.41 per cent), Year 5 (11.62 per cent), and Year 6 (13.04 per cent). Examples of words at CEFR A2 are collects, explain, lay, medicine, and temperature. The words are common and slightly less familiar as words in CEFR A1 level.

The analysis in all the six excerpts from the different grade level showed that the percentage of CEFR A1 and A2 in each excerpt is significant: Year 1 (62.79 per cent), Year 2 (60.11 per cent), Year 3 (70.75 per cent), Year 4 (55.94 percent), Year 5 (50 per cent) and Year 6 (49.83 per cent). Despite the high percentage of words at CEFR A1 and A2 levels, the analysis from the Scorecard showed that all the excerpts are beyond CEFR A2 level of difficulty. This is in view of the significant number of words in BI, B2, C1 and *unlisted*. *Unlisted* word refers to words that are not listed in the English Vocabulary Profile according to the level of the CEFR, which indicates the least used words, terminology, proper noun, and numbers.

CEFR B1 Words
The analysis of words from CEFR B1 indicates the difficulty of the texts; however, the difficulty of the words at this level is believed still manageable to be addressed by the teachers (see Table 5).

Table 5

Example of words and percentage at CEFR B1

Year	B1 (%)	Example of Words
1	9.30	butterfly, frog, importance, protect, similar, skin, wing, wings
2	14.45	ants, based, branches, butterfly, calf, cycle, dolphin, dolphins, feed, frog, frogs, fur, ground, hidden, kangaroo, kittens, look like, meanwhile, mosquito, penguins, protect, similar, therefore, thick, tigers
3	10.88	according to, based, compare, goat, habits, mentioned, natural, regions, seeds, sharp, sharper, similar, situations, tear, therefore, tigers
4	10.89	according to, although, based, bee, besides, breathe, breathing, butterfly, climates, dolphins, frog, frogs, land, method, methods, mosquito, mostly, normally, regions, skin whales, wonder
5	16.16	ability, attack, attacked, avoid, based, bees, behaviours, certain, defence, enemies, enemy, escape, fortunately, frogs, ink, patterns, produce, protect, protected, raise, reach, sharp, similar, situation, skin, smell, stick, thick, touch, towards, used to, wings
6	11.71	actions, because of, breaking down, cheek, conditions, damage, deaths, disease, diseases, due, effects, equipment, exist, experiment, fight, flu, human, humans, mainly, objects, packets, prevents, produce, producing, products, proving, skin, spoil, suitable, therefore, tiny, used to, virus, viruses, waste

Based on the sample of the CEFR B1 words, the words can be difficult for students who do not have the proficiency at least at CEFR A2 level. However, these words should be still manageable by science teachers and they should be able to explain these words as the words are still familiar and some are common in everyday contexts such as ability, damage, natural, protect, and virus.

CEFR B2 words

The analysis below showed that CEFR B2 words are significantly used in textbooks at all levels (see Table 6).

Table 6

Example of words and percentage at CEFR B2

Year	B2(%)	Example of Words
1	13.95	crocodile, feather, function, horn, horns, identifies, protection, scale, shell, similarities, steering, tail
2	7.51	crocodile, features, give birth, giving birth, growth, lay eggs, laying eggs, newly, observe, observing, occur, state, swans
3	4.76	contrast, eagle, naturally, observe, polar bears, swallowing, worms
4	11.39	characteristics, compared to, crab, crocodile, deer, feathers, give birth, lay eggs, lays eggs, live on, living on, lungs, observe, pond, scales, shell, specific, structures, surface, tropical, unique, worm, worms
5	14.65	causing, characteristics, crab, crocodiles, deceive, distract, eagle, ensure, fail to, horns, identify, injuries, maintain, muscles, observe, poisonous, protection, release, run away, run into, scales, shell, shells, species, specific, survival, threatened, threats, wasps
6	12.71	absence, absorbed, absorbing, acid, affecting, cause, caused, causes, characteristics, combination, decay, disrupt, factors, growth, harmful, heard of, identified, infection, infections, live on, manufacture, manufacturing, naked, occur, occurs, organic, poisoning, ponds, presence, process, production, soil, sources, state, stimulate, stored, survive, treatment

The analysis indicates that even though the words are classified as CEFR B2, the words from the textbooks are still familiar in everyday context (e.g., crocodile, eagle, lay eggs, soil, worm) provided that students are at CEFR B1. At single word level, the CEFR B2 words still can be perceived familiar at least to students of CEFR B1 level and teachers (assuming teachers are at CEFR B2). However, the contents of science subject tend to be abstract and not a familiar scenario. Partial understanding of the words as well as lack of experience/imagination of the scenario may easily affect comprehension. This analysis indicates that teachers who have the proficiency of CEFR B2 may not find that these CEFR B2 words are difficult to understand; however, students of CEFR A2 proficiency may have difficulty to understand the science texts.

CEFR C1 words
The analysis shows that percentage of words at CEFR C1 is small but enough to impede comprehension (see Table 7).

Table 7
Example of words and percentage at CEFR C1

Year	C1 (%)	Example of Words
1	2.33	beak, self
2	2.31	laid, reproduce, reproduced, reproduction
3	2.72	classified, classify, happen to, mustard
4	5.45	classified, classify, conclude, habitat, habitats, mammal, mammals, organ, organs, reproduction, reptiles
5	2.53	extinction, fake, resemble, self, spines
6	1.34	bacteria, infected, makes it, reproduce

Table 7 shows that CEFR C1 words are used in Years 1 to 6 of science textbook. Even there is less than 6% of CEFR C1 words are being used in the textbooks, these words might hamper students' comprehension in learning. The extracted words are words commonly used in the field of science, not words commonly used in everyday context. At CEFR C1, most of the words are more specialised to the field.

CEFR C2 words
Below is the example of words at the CEFR C2 level that are available in the science textbooks except in Year 1 textbook (see Table 8).

Table 8
Example of words and percentage at CEFR C2

Year	C2 (%)	Example of Words
1	-	-
2	0.58	Nests
3	1.36	classification, classifications
4	0.99	classification, offspring
5	1.01	prey, scare the
6	1.67	antibiotic, infectious, particles, spiral, vaccines

From the analysis, there is less than 2% of the CEFR C2 words appear in the science textbooks. However, those words are still crucial in science context. Therefore, science teachers need to be aware that CEFR C2 words might need further explanation in the learning process.

Unlisted Words

In all the samples analysed, the percentages of *Unlisted* words (see Table 9). The *Unlisted* words are words that contain fewer than the minimum number of raw occurrences for a CEFR level (Capel, 2012). As such, these words, which include proper nouns (names and places), numbers, abbreviations, and specific-domain words, are not listed in the EVP. Therefore, these words are not included in the EVP, which includes proper nouns, numbers, abbreviations, and specific-domain words.

Table 9
Example of words and percentage of Unlisted

Year	Unlisted (%)	Example of Words
1	11.63	antenna, dragonfly, fin, fins, Hafiz, hamster, rhinoceros, snail, tortoise, webbed
2	15.03	's, animals', caterpillar, chicks, cubs, cute, froglet, grasshoppers, hatched, hatching, hippopotamuses, Kugan, ladybird, ladybirds, larva, larvae, moth, moths, Nisa, ostrich, pangolins, porcupine, stork, storks, tadpole, yearling
3	9.52	's, ah, animals', argh, canines, carnivores, dentition, grind, herbivores, incisors, molars, omnivores, orangutans, polar
4	15.35	amphibians, backbone, backbones, caecilian, caecilian, caecilias, caterpillar, coverings, dragonfly, earthworm, freshwater, gills, grasshopper, grasshoppers, invertebrate, invertebrates, leech, moist, newts, offsprings, platypus, platypuses, salamenders, spiracles, stork, suckles, tadpole, temperate, tortoise, vertebrate, vertebrates
5	15.66	's, 0, 30 arapaima, bedbugs, beetles, buffaloes, burrows, centipedes, claws, cockroaches, cuttlefish, deers, detach, gaurs, horseshoe, inflate, lizards, moths, octopuses, pierced, porcupines, porcupines', prickly, pufferfish, quills, rhinoceros, spurt, stings, tortoise, venom
6	22.74	19, 2019, acidity, algae, amoeba, antibodies, bacteriophage, bacterium, China, Chlamydomonas, chlorella, chlorophyll, coli, contaminated, coronavirus, cov, covid, decomposed, decomposing, decomposition, electron, Escherichia, faeces, fermented, fertiliser, fertilisers, fluffy, fungi, fungus, Hadi, hiv, immunodeficiency, inactive, influenza, measles, micro, microorganism, microorganisms, microscopes, moisture, mould, mumps, nutrients, ones, organism, paramecium, penicillin, penicillium, photosynthesis, protozoa, respiratory, Rhizopus, ringworm, rod, salmonella, sawdust, sewage, shown, specialised, spherical, spirillum, spoilage, spores, streptococcus, swelling, tempeh, volvox, yeast

The analysis indicates that proper nouns such as name (Hafiz, Kugan) and name of animals (rhinoceros, caterpillar) are listed as *Unlisted* words even though they do not have

meaning. A significant percentage (10 to 20 per cent) of the *Unlisted* words comes from science words such as photosynthesis, penicillin, spoilage, and others. Based on the percentages, the *Unlisted* words can be perceived unavoidable, as the words are not easily replaced with other words or simpler words, and the synonyms may affect meaning.

Conclusion

In conclusion, this study found that the language used in science textbooks from Year 1 to Year 6 is at CEFR B1 to C1, in which beyond CEFR A2 level. The analysis also revealed that the percentage of words beyond CEFR A2 level is between 30 to 50 percent. With this level of language difficulty, it is questionable that students can understand the textbooks without assistance. This finding substantiates past studies on the readability of learning materials (see e.g., Ayodele, 2013; Francis et al., 2020).

The meta-analysis of words indicates that even though the words are of CEFR B1 and B2, the words are not complicated and can easily be managed by teachers, subject to the teacher's awareness on the difficulty of the words. It is important to note that learning science in primary schools is mostly text-based, not experimental. Thus, understanding the content relies on the comprehension of the content in the textbook. Given the current level of language used in the textbooks, students need to be guided by teachers to understand the textbooks. As Seah and Silver (2018) found, teachers must pay attention to students' language demands, especially in teaching scientific vocabulary.

The findings on the CEFR C1 and C2 words in the samples analysed are unexpected. These words are found in all levels of the textbooks. They are science-content words. To those who have adequate proficiency such as science teachers, the words (such as extinction, reproduction, particles) may not be considered difficult to understand; hence teachers may not be aware of the need to explain them. Thus, this study intends to document the wordlist of the DLP science textbooks to guide science teachers, teaching material writers and assessment writers to be aware of the suitable words according to students' level of English proficiency.

The wordlist is beneficial to create the awareness and help science teachers to identify words that may pose difficulty for students to comprehend the lesson as well as in designing assessment tasks, and to those involved in textbook/learning material writing. Last but not least, the wordlist provides a starting point to consider a more extensive research in the teaching and learning materials when English is not the dominant language of students and teachers.

Acknowledgement

References

Ali, M., & Ismail, Z. (2006). Comprehension level of non-technical terms in science: Are we ready for science in English. *Malaysian Journal of Educators and Education*, *21*, 73-83.

Ayodele, M. O. (2013). A comparative study of textbook readability and students' comprehension levels in senior secondary school biology. *Journal of Educational and Social Research*, *3*(1), 109. https://doi.org/10.5901/jesr.2013.v3n1p109

Capel, A. (2012). Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal*, *3*(1), 1–14. https://doi.org/10.1017/s2041536212000013

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment.* Council of Europe Publishing, Strasbourg. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Francis, N. K., Eric, A. A., PeprahMensah, D., Irene E. N., & AbaBrew-Hammond. (2020). How can we take up science courses? Readability of levels of junior high school integrated science textbook in Ghana. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, *25*(8), 16–24. https://doi.org/10.9790/0837-2508061624

Gyasi, W. K. (2013). The role of readability in science education in Ghana: A readability index analysis of Ghana association of science teachers textbooks for senior high school. *IOSR Journal of Research & Method in Education (IOSR-JRME), 2*(1), 9–19. https://doi.org/10.9790/7388-0210919

Masrom, U. K., Idris, M., & Jusoh, Z. (2021). Dual Language Programme (DLP): Mediating effects of readiness, interest and confidence on students' satisfaction. *Journal of Nusantara Studies (JONUS)*, *6*(2), 1–20. https://doi.org/10.24200/jonus.vol6iss2pp1-20

Ministry of Education Malaysia. (2014). *Dasar memartabatkan Bahasa Malaysia memperkukuhkan Bahasa Inggeris: MBMMBI*. Kementerian Pendidikan Malaysia.

Ministry of Education Malaysia. (2015). *English language education reform in Malaysia the roadmap 2015-2025*. http://eltc.moe.edu.my/roadmap/The Roadmap 2015-2025.pdf

Ministry of Education Malaysia. (2020). *Surat pekeliling ikhtisas kementerian pendidikan malaysia bilangan 3 tahun 2020 Penambahbaikan garis panduan pelaksanaan Dual language programme (DLP) kementerian pendidikan Malaysia tahun 2021*. https://www.moe.gov.my/pekeliling/3873-spi-bilangan-3-tahun-2020-penambahbaikan-garis-panduan-pelaksanaan-dual-language-programme-kementerian-pendidikan-malaysia-tahun-2021/file

Don, M. Z., & Abdullah, H. (2019). The reform of English language education in Malaysia. *FMT*. https://www.freemalaysiatoday.com/category/opinion/2019/05/22/the-reform-of-english-language-education-in-malaysia/

Readable. (2011). https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/

Seah, L. H., & Silver, R. E. (2018). Attending to science language demands in multilingual classrooms: A case study. *International Journal of Science Education*, 2453–2471. https://doi.org/10.1080/09500693.2018.1504177

Text Inspector (n.d.). https://textinspector.com/

Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies*, *2*(1), 43–53. https://doi.org/10.4304/tpls.2.1.43-53