A Corpus-Based Analysis of Vocabulary Progression in Philippines' English Modules

Mary Darlene Q. Burgos

Department of Education, Philippine Normal University

DOI: https://doi.org/10.47772/IJRISS.2026.1026EDU0286

Received: 11 May 2026; Accepted: 16 May 2026; Published: 4 June 2026

ABSTRACT

The present study investigated vocabulary progression within the Philippine K-12 English curriculum by conducting a corpus-based analysis of selected Grade 10 and Grade 11, the crucial transitioning years between junior and senior high school. This research compared the lexical profiles of the said modules and determined if Grade 11 learning materials that are generally aligned to English as a subject exhibits increased lexical complexity, aligning with the principles of a spiraling curriculum. A quantitative, comparative corpus linguistics methodology was employed, utilizing AntConc and #LancsBox software to analyze metrics including token/type counts, Type-Token Ratio (TTR), Moving-Average TTR (MATTR), Measure of Textual Lexical Diversity (MTLD), keyword frequency and likelihood, dispersion (Julliand's D), and collocates. Findings revealed that the Grade 11 corpus was significantly larger and contained more unique word types than the Grade 10 corpus. While the results show that Grade 10 has higher TTR than Grade 11, MATTR indicated similar local lexical diversity. Notably, numbers suggest that Grade 11 has greater overall lexical diversity (MLTD). Keyword, dispersion, and collocate analyses indicated nuanced shifts in subject matter and contextual word usage, moving from foundational concepts and tasks in Grade 10 towards more analytical and specialized applications in Grade 11. The results provide empirical evidence supporting vocabulary progression and increased complexity from Grade 10 to Grade 11, largely aligning with K-12 expectations. These findings imply that while the modules generally reflect the expected rise in difficulty across the K-12 curriculum, this progression is characterized by more than just vocabulary growth. It also includes variations in subject matter focus and the practical use of language.

Keywords: Corpus, Modules, Vocabulary, Spiral Progression

INTRODUCTION

Rationale of the Study and its Significance

According to Cameron (2001), the design of language learning materials should ensure that learners have opportunities to see new vocabulary frequently and in a range of different settings. In order to acquire language, one must be exposed to materials that help increase one’s vocabulary.

Four years have passed since the Philippine education system underwent a rapid paradigm shift to ensure learning continuity during the pandemic. This transition significantly altered teaching and learning approaches in the country, embracing online and modular learning as alternatives to traditional face-to-face instruction. The schools opted for modules as a temporary solution, offering self-contained learning packages with content, activities, and assessments (Bustillo & Aguilos, 2022). In the Department of Education (DepEd) Order No. 009, s. 2024, "Implementing Guidelines on the School Calendar and Activities for the School Year 2024-2025," modular learning is defined as “self-directed learning" facilitated by learning modules. Despite the return to face-to-face classes for the past three school years, DepEd Division Offices continue to produce these modules as guides for both teachers and learners. According to Ambayon (2020), modular approach help maximize the students’ involvement especially if teachers obtain the adequate training to strategize and design the module for instruction.

There are several researches that used corpus study to investigate the content of textbooks and learning materials but very little in the context of Philippines. The researches from Costales (2022) and Gumangan and Dita (2022) focused on corpus produced by academic works written by students. Thus, further research is needed to look through modules that were used under K-12 program in the Philippines.

The present study aims to address this gap by conducting a corpus-based analysis of selected official Grade 10 and Grade 11 English learning modules utilized within the Philippine K-12 system. The primary objective is to empirically examine and compare the vocabulary profiles of these materials, focusing specifically on aspects related to lexical progression, recycling, and complexity.

Furthermore, the analysis seeks to determine whether the Grade 11 modules present a demonstrable increase in lexical complexity compared to their Grade 10 counterparts. To achieve these aims, this research employs methodologies drawn from corpus linguistics, involving the compilation and computational analysis of corpora derived directly from the selected SLMs. The findings are expected to provide valuable data on the nature of vocabulary progression in these educational resources, offering insights for curriculum planners, materials developers, and English language educators within the Philippines and potentially contributing to the broader field of second language (L2) materials analysis.

LITERATURE REVIEW

To date, there has been relatively few corpus linguistic research specifically examining the development of textbooks and their vocabulary from a teaching perspective especially in the Philippines. Most corpus studies would opt for written outputs of learners which would usually suggest strategies and approaches in teaching L2(eg. Costales, 2022; Gumangan & Dita, 2022).

Although textbooks and learning modules are crucial in English language learning, little research has investigated how well they aid language acquisition and the relevance of their vocabulary content. One of those researches is Norberg and Nordlund (2018) who examined the vocabulary of seven English textbooks utilized in Swedish primary schools.

They also used corpus analysis, facilitated by concordance software, the study compared the lexical content across these textbooks and against established word lists (the New General Service List and VP-Kids corpus). The findings revealed a significant presence of low-frequency words in everyday language and substantial inconsistencies in the quantity and selection of vocabulary across the textbooks, suggesting a lack of a unified approach to word choice in these educational materials.

In China, a research conducted by Yu and Renandya (2021) analyzed the vocabulary of the widely-used New Senior English for China textbooks against China's 2017 curriculum. They have found that the textbooks cover only about 80% of the required words, with half appearing infrequently. While frequently repeated words are generally well-distributed, poorly dispersed words are mostly thematic. Overall, the textbooks lack sufficient vocabulary coverage and repetition, necessitating extra learning.

A study in Taiwan by Hsu (2009) built a vocabulary database of General Education (GE) textbooks in universities in Taiwan to analyze their size, frequency levels (using the British National Corpus), and text coverage, including Coxhead's Academic Word List (AWL) and General English Proficiency Test vocabulary. The results indicated that GE textbooks offer 49-415 academic words and 162-2,001 new general words, covering 24.55%-65% of intermediate GEPT vocabulary.

A research published from Rahva Journal of Technical and Social Studies by Söğüt and Vural (2023) conducted the study that compared the vocabulary of a "corpus-informed" and a conventional pre-intermediate English textbook. The research revealed a surprising result: the conventional book covered more of the top 2000 English words. Further analysis showed no clear advantage in vocabulary variety or repetition for the "corpus-informed" book, suggesting its lexical choices weren't systematically driven by frequency. These international studies underscore the need for more research into textbook vocabulary to better support language acquisition.

Statement of the Problem

This research is guided by the following questions:

1. What are the overall statistical profiles of the vocabulary used in the selected Philippine Grade 10 and Grade 11 learning modules?
2. To what extent is core vocabulary from the Grade 10 module reintroduced, reused, or built upon in the Grade 11 module?
3. Do the observed patterns of vocabulary introduction, distribution, repetition, and increasing complexity between the Grade 10 and Grade 11 modules align with the expected characteristics K-12 Curriculum?

METHODOLOGY

This study employed a quantitative, comparative corpus linguistics approach within descriptive design to analyze the distribution and complexity of vocabulary in selected Grade 10 and Grade 11 learning modules from the Philippines.

Research Design

The core design is a comparative analysis of two purpose-built corpora, one representing a Grade 10 English learning modules and the other a corresponded with Grade 11 English Subjects namely: Reading and Writing Skills, 21st Century Literature from the Philippines and the World, and Oral Communication. Statistical measures of vocabulary frequency, diversity, density, dispersion, and collocation will be used to address the research questions and test the hypotheses.

Data Collection

Adopting a descriptive design, this study utilizes quantitative comparative corpus linguistics to analyze whether the modules issued for Grade 10 (Quarter 1-4), and the modules for Senior High School Subjects usually offered in Grade 11. The modules for this corpus study that represent Grade 11 subjects were Reading and Writing Skills, 21st Century Literature from the Philippines and the World, and Oral Communication. The modules for Grade 11 hypothetically adhere to spiral progression in terms of vocabulary skills in connection with Grade 10 modules.

This is to note that each Schools Division of DepEd may have different content but follows the same set of topics. In this case, this research used the Modules issued in Schools Division Office of Valenzuela as corpus.

There are four sets of modules in pdf format in Grade 10 modules. The reference corpus which is the Grade 11 modules has a total of six files (two .pdf files per subject). All corpora have been converted into .txt files for the computer software to run the corpus data smoothly.

Data Analysis

The sample corpus was run through Ant Conc which is a freeware software program for working with language corpora using a graphical user interface. Alongside the latter, is an application called #LancsBox with is also a software package developed at Lancaster University for the analysis of language data and corpora.

To determine the overall statistical profile of the chosen corpus, the keywords, likelihood, and collocates, the researcher used AntConc. To determine the dispersion of words throughout the corpus, #LancsBox has helped in determining the Julliand's D of the keywords in corpus.

Data Gathering Procedure

To systematically gather and analyze the corpus data for this research, the following procedure was implemented. First, the Grade 10 modules, comprising four sets of materials available in PDF format, were identified and collected. Subsequently, a reference corpus consisting of Grade 11 modules was compiled, totaling six files

organized into two PDF files per subject. To facilitate computational analysis, both the Grade 10 modules and the Grade 11 reference corpus were converted from their original PDF format into plain text (.txt) files. This conversion ensured smooth processing and compatibility with the chosen corpus analysis software. Following the conversion, AntConc software was utilized to determine the overall statistical profile of the Grade 10 corpus, specifically focusing on identifying key words, calculating likelihood scores, and extracting collocates. Finally, to investigate the distribution of specific keywords throughout the Grade 10 corpus, #LancsBox software was employed to calculate Julliand's D, providing insights into the dispersion patterns of these terms within the textual data.

Ethical Considerations

This research focuses on the analysis of publicly accessible educational materials, specifically DepEd modules released by the Schools Division Office of Valenzuela. As such, it does not directly involve human participants, mitigating ethical concerns related to informed consent, privacy, and potential harm to individuals. While the materials are publicly available for educational purposes, they are still the intellectual property of the Department of Education and the individuals who created them. This research adheres to the highest standards of academic integrity by ensuring meticulous citation and clear acknowledgement of the source materials in all stages of the research and dissemination. This includes not only direct quotations but also the overall structure and content of the modules if they significantly influence the analysis.

The analysis of these educational materials is conducted with rigor and objectivity to avoid misinterpretations of their content or pedagogical intent. The research strives for a fair and accurate representation of the vocabulary, language use, and potential learning opportunities embedded within the modules. This involved careful contextualization of findings and avoiding generalizations beyond the scope of the analyzed materials.

Although no direct human participants are involved, the findings of this research have the potential to indirectly influence educational practices, curriculum development, and the perceptions of the quality of these learning resources. Therefore, the research was conducted with a sense of responsibility for its potential impact. The dissemination of findings will be framed constructively, focusing on providing insights that could contribute to the improvement of educational materials and language teaching practices in the Philippines. As the research focus on materials used in the Philippine educational system, particularly in Valenzuela, it is important to be mindful of the local context and cultural sensitivities. The analysis will avoid making value judgments based on external standards without considering the specific educational goals and cultural nuances of the Philippine curriculum.

RESULTS AND DISCUSSION

Having analyzed the sample corpus with AntConc and #LancsBox to establish its statistical properties, including keywords, likelihood, collocates, and word dispersion (Julliand's D), the subsequent section offers a general comparison of the entire Grade 10 and Grade 11 module collections.

Table 1: Overall Statistical Profile Comparison of Grade 10 and Grade 11 Modules

Metric	Grade 10 Module	Grade 11 Module	Difference (G11 - G10)	Ratio (G11 / G10)
Corpus Size				
Total Word Tokens	81598	99080	17,482	1.21
Total Word Types (Unique Words)	7293	8051	758	1.10
Vocabulary Richness/Diversity				
Simple Type-Token Ratio (TTR) (%)	0.089	0.081	-0.81%	0.91

The Grade 11 module has a significantly larger corpus size with 99,080 total word tokens compared to the Grade 10 module's 81,598. The difference is 17,482 more tokens in Grade 11, representing a ratio of 1.21 (Grade 11 is 1.21 times larger than Grade 10 in terms of total words).

Similarly, the Grade 11 module contains more unique words (8,051) than the Grade 10 module (7,293). The difference is 758 unique words, with a ratio of 1.10 (Grade 11 has 1.10 times more unique words than Grade 10).

In terms of vocabulary richness or diversity, the Simple Type-Token Ratio (TTR) indicates the proportion of unique words to the total number of words, providing a basic measure of lexical diversity. The Grade 11 module has a slightly lower TTR (0.081 or 8.1%) compared to the Grade 10 module (0.089 or 8.9%). The difference is -0.81%, and the ratio is 0.91, suggesting that while Grade 11 has more unique words, its overall vocabulary richness (in terms of the proportion of unique words) is slightly lower than Grade 10. This could imply more repetition of words in the Grade 11 module.

Table 2: Table 2: Moving-Average Type-Token Ratio and MTLD

	Tokens	MATTR	MTLD
Grade 10 Module 1	25465	0.78	72.2
Grade 10 Module 2	21570	0.78	72.53
Grade 10 Module 3	20687	0.78	69.83
Grade 10 Module 4	15152	0.78	70.23
Grade 11 OralCom Q1	9797	0.78	74.07
Grade 11 Oralcom Q2	9047	0.78	74.07
Grade 11 RAW Q1	13016	0.78	74.07
Grade 11 RAW Q2	12086	0.81	74.07
Grade 11 21st Q1	24824	0.78	64.87
Grade 11 21st Q2	18439	0.79	76.47

In this table, the MATTR (Moving-Average Type-Token Ratio) metric calculates TTR over a fixed-size moving window of tokens and then averages these TTR scores. It is less sensitive to text length than the simple TTR. Across all listed modules (both Grade 10 and Grade 11), the MATTR values are relatively consistent, ranging from 0.78 to 0.81. This suggests a similar level of lexical diversity when considering moving windows of text within each module.

The MTLD (Measure of Textual Lexical Diversity) calculates the average length of word sequences required to maintain a certain TTR threshold. Higher MTLD values generally indicate greater lexical diversity. The MTLD values show a clear distinction between Grade 10 and Grade 11 modules. Grade 10 modules have MTLD values ranging from approximately 69.83 to 72.53. Grade 11 modules consistently show higher MTLD values, ranging from 74.07 to 76.47 (with one exception at 64.87 for Grade 11 21st Q1).

As expected, the Grade 11 modules are larger in terms of both total word count and the number of unique words. This likely reflects the increasing complexity and scope of the curriculum. Despite having more unique words, the slightly lower simple TTR in Grade 11 suggests a potentially higher rate of word repetition compared to Grade 10. The relatively similar MATTR values indicate that the local lexical diversity within comparable segments of text is quite similar between the two grade levels.

The consistently higher MTLD scores for most Grade 11 modules suggest a greater overall lexical diversity in these modules compared to Grade 10. This metric is less influenced by text length and indicates that Grade 11 texts can sustain a higher proportion of unique words over longer stretches of text. The Grade 11 21st Q1 module

has a notably lower MTLD (64.87) compared to other Grade 11 modules. This might warrant further investigation to understand the specific characteristics of this module's vocabulary.

In summary, the Table 1 and Table 2 indicates that while Grade 11 modules are larger and contain more unique words, their overall lexical diversity, as measured by MTLD, tends to be higher than that of Grade 10 modules. The simple TTR suggests a slightly higher repetition rate in Grade 11, but the more robust MATTR shows similar local diversity. The outlier in the Grade 11 21st Q1 MTLD value is worth noting.

Table 3: Top 20 Keyword Frequency and Likelihood

Type	G10 Frequency	G11 Frequency	G10 NormFreq	G11 NormFreq	G10 NormRange	G11 NormRange	Keyness (Likelihood)	Keyness (Effect)
is	1327	1159	16282.61	13450.47	1	1	22.993	0.032
an	406	297	4981.717	3446.755	1	1	23.687	0.01
english	405	88	4969.447	1021.261	1	1	239.602	0.01
research	240	39	2944.858	452.604	1	0.833	172.701	0.006
information	238	121	2920.317	1404.234	1	1	45.731	0.006
used	222	106	2723.993	1230.155	1	1	48.737	0.005
should	174	102	2135.022	1183.734	1	1	23.266	0.004
statement	157	81	1926.428	940.024	1	1	29.159	0.004
listening	142	7	1742.374	81.237	0.5	0.667	157.833	0.003
I	136	28	1668.753	324.947	1	1	83.662	0.003
help	129	70	1582.861	812.367	1	1	21.225	0.003
covid	124	27	1521.51	313.341	1	0.833	73.1	0.003
claim	122	54	1496.969	626.683	1	1	30.926	0.003
definition	100	23	1227.024	266.92	1	0.667	56.408	0.002
persuasive	89	26	1092.051	301.736	1	0.5	40.114	0.002
argumentative	85	4	1042.97	46.421	1	0.167	95.368	0.002
data	85	12	1042.97	139.263	1	0.5	66.042	0.002
study	76	16	932.538	185.684	1	0.833	45.962	0.002
terms	75	18	920.268	208.894	1	0.833	40.808	0.002
article	71	25	871.187	290.131	1	0.5	25.626	0.002

The specific word frequencies reveal interesting patterns between the Grade 10 and Grade 11 modules. Common function words like "is" and "an" exhibit high frequencies in both corpora with normalized frequency ratios close to one and low keyness and effect values, indicating similar relative usage across both grade levels.

However, content-specific words show a different trend. For instance, "english" appears significantly more frequently in Grade 10, as indicated by its low normalized frequency ratio and high positive keyness, suggesting a greater emphasis on the subject itself in the Grade 10 materials analyzed. Similarly, terms like "research," "information," "used," "should," "statement," "listening," "I," "help," "covid," "claim," "definition," "persuasive," "argumentative," "data," "study," "terms," and "article" all display higher frequencies and normalized frequencies in the Grade 10 module, accompanied by positive keyness and effect values.

This collective trend suggests that the Grade 10 content in this comparison places a greater relative emphasis on these specific vocabulary items, potentially reflecting differences in subject matter, learning activities, or writing styles compared to the Grade 11 module.

Notably, the finding that "research" is more key to Grade 10 in this specific word analysis contrasts with the broader corpus analysis from the previous table. This discrepancy underscores the critical importance of pairing macro-level corpus analysis with micro-level individual word examinations. Ultimately, these findings imply

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

that curriculum progression is not merely a linear increase in difficulty, but a complex shifting of thematic priorities and contextual language applications across transitioning academic levels.

is	None	None	1	1327	1159	4	6	16282.61
an	None	None	4	406	297	4	6	4981.717
english	None	None	5	405	88	4	6	4969.447
research	None	None	6	240	39	4	5	2944.858
information	None	None	7	238	121	4	6	2920.317
used	None	None	8	222	106	4	6	2723.993
should	None	None	9	174	102	4	6	2135.022
statement	None	None	10	157	81	4	6	1926.428
listening	None	None	12	142	7	2	4	1742.374
l	None	None	13	136	28	4	6	1668.753
help	None	None	14	129	70	4	6	1582.861
covid	None	None	15	124	27	4	5	1521.51
claim	None	None	16	122	54	4	6	1496.969
definition	None	None	17	100	23	4	4	1227.024
persuasive	None	None	18	89	26	4	3	1092.051
argumentative	None	None	19	85	4	4	1	1042.97
data	None	None	19	85	12	4	3	1042.97
study	None	None	21	76	16	4	5	932.538
terms	None	None	22	75	18	4	5	920.268
article	None	None	23	71	25	4	3	871.187
technical	None	None	24	70	1	3	1	858.917
speaking	None	None	25	65	16	3	6	797.566
opinion	None	None	25	65	19	4	4	797.566
chart	None	None	27	64	14	4	5	785.295
negation	None	None	28	61	0	2	0	748.485
health	None	None	28	61	14	4	4	748.485
support	None	None	28	61	20	4	5	748.485
sources	None	None	31	60	13	4	5	736.214
title	None	None	32	57	15	4	6	699.404
affirmation	None	None	33	50	0	2	0	613.512
pandemic	None	None	33	50	9	4	5	613.512
news	None	None	33	50	13	4	4	613.512
operational	None	None	36	49	0	1	0	601.242
cohesion	None	None	37	48	4	3	1	588.972
primary	None	None	39	44	12	4	4	539.891
issue	None	None	40	40	10	4	4	490.81
spread	None	None	41	37	2	3	2	453.999
feminist	None	None	41	37	5	2	1	453.999
definitions	None	None	43	36	1	4	1	441.729
structuralism	None	None	44	34	0	1	0	417.188

local	None	None	44	34	6	4	1	417.188
moralist	None	None	46	33	0	2	0	404.918
timeline	None	None	46	33	1	2	1	404.918
listener	None	None	46	33	2	2	1	404.918
verb	None	None	46	33	3	4	3	404.918
bibliography	None	None	50	32	0	2	0	392.648
website	None	None	50	32	5	3	1	392.648
theory	None	None	52	31	6	2	4	380.377
physical	None	None	53	30	5	3	2	368.107
assertion	None	None	54	29	0	1	0	355.837
secondary	None	None	54	29	2	2	1	355.837
citation	None	None	56	28	0	3	0	343.567
cohesive	None	None	56	28	0	3	0	343.567
factual	None	None	56	28	3	3	1	343.567
apa	None	None	59	27	0	1	0	331.296
publication	None	None	60	26	3	3	2	319.026
formalism	None	None	61	25	0	1	0	306.756
infection	None	None	61	25	3	4	1	306.756
organizers	None	None	63	24	1	2	1	294.486
skill	None	None	63	24	2	4	2	294.486
marxist	None	None	65	23	0	2	0	282.216
singular	None	None	66	21	0	1	0	257.675
feminism	None	None	67	20	1	1	1	245.405
marxism	None	None	69	18	0	1	0	220.864
germs	None	None	70	17	0	2	0	208.594
episode	None	None	70	17	0	1	0	208.594
mla	None	None	70	17	0	1	0	208.594
shock	None	None	73	16	0	2	0	196.324
karl	None	None	73	16	0	4	0	196.324
healthy	None	None	73	16	0	1	0	196.324
structuralist	None	None	76	15	0	2	0	184.054
citations	None	None	76	15	0	2	0	184.054
instructional	None	None	79	14	0	2	0	171.783
workers	None	None	79	14	0	4	0	171.783

Table 4: Term Frequency and Julliand's D by Grade Level

	Grade 10			Grade 11		
Terms	Frequency	Range	Juliand's D	Frequency	Range	Juliand's D
english	61	100	0.98	32	100	0.85
research	237	100	0.39	41	85.71	0.52
information	238	100	0.67	155	100	0.76
used	221	100	0.8	136	100	0.77
should	178	100	0.8	113	100	0.8

statement	157	100	0.65	88	100	0.85
listening	140	50	0.01	10	71.43	0.65
help	126	100	0.86	80	100	0.92
covid	22	75	0.53	4	57.14	0.64
claim	121	100	0.41	55	100	0.17
definition	98	100	0.05	38	71.43	0.5
persuasive	87	100	0.26	29	57.14	0.23
argumentative	80	100	0.38	8	28.57	0.35
data	84	100	0.49	14	57.14	0.55
study	73	100	0.51	17	85.71	0.73
terms	74	100	0.45	22	85.71	0.76
article	68	100	0.5	25	57.14	0.28
technical	69	75	0.06	1	14.29	0
speaking	58	75	0.2	17	100	0.53

Julliand's D offers a crucial perspective on how the terms are distributed. Terms like "english" (G10: 0.98, G11: 0.85), "used" (G10: 0.8, G11: 0.8), and "help" (G10: 0.86, G11: 0.92) tend to be more evenly distributed throughout the respective grade level's materials. In contrast, terms like "listening" (G10: 0.01, G11: 0.65), "covid" (G10: 0.53, G11: 0.64), "definition" (G10: 0.05, G11: 0.5), "technical" (G10: 0.06, G11: 0), and "argumentative" (G10: 0.38, G11: 0.35) appear to be more clustered or concentrated in specific sections of the curriculum for that grade level. A Julliand's D of 0 for "technical" in Grade 11 suggests it might appear in only very limited or specific contexts.

Some terms show notable changes in their dispersion from Grade 10 to Grade 11. For example, "listening" has a very low Julliand's Din Grade 10 but a moderate one in Grade 11, suggesting it might be a focused topic in specific Grade 10 materials but more integrated throughout the Grade 11 curriculum. Similarly, "definition" and "technical" show increased dispersion in Grade 11 (though "technical" remains very low).

Table 4 provides a valuable perspective on the distribution of key terms within the Grade 10 and Grade 11 learning materials. By examining both the frequency and Julliand's D, we can gain a deeper understanding of how these terms are integrated into the curriculum, whether they are core concepts consistently used or more localized vocabulary within specific topics or activities. The differences in frequency and dispersion between the two grade levels can highlight shifts in curricular focus and the way specific concepts are addressed as students' progress.

Broadly speaking, these variations in Julliand's D values provide empirical evidence of how a spiraling curriculum organizes knowledge during a critical academic transition. The data demonstrates that while core structural vocabulary (like "english" or "help") remains stably and evenly dispersed across both years, specialized academic terms undergo a structural shift.

Moving from highly clustered, low-dispersion values in Grade 10 to higher dispersion in Grade 11 suggests that concepts once taught in isolated, dedicated units (such as "listening" or "definition") eventually become integrated, foundational tools used throughout the later curriculum.

For curriculum designers and educators, this implies that student progression into senior high school is not just about encountering entirely new words, but about encountering familiar words more frequently and across a wider, more diverse range of learning contexts.

Ultimately, this highlights that vocabulary mastery in a transitioning curriculum is as much about structural integration as it is about simple exposure.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

Table 5: Top 10 Keywords and their Collocates in Grade 10 and Grade 11

Core Term	Top 3 Collocates G10	Grade 10 Likelihood	Top 3 Collocates G11	Grade 10 Likelihood
english	do	1448.164	lesson	265.738
	module	1191.264	module	225.996
	lesson	864.854	do	214.739
research	paper	87.407	review	43.08
	methodology	72.236	journal	24.949
	terms	62.251	step	23.893
information	provide	110.793	dissemination	120.202
	literacy	80.193	organizing	105.301
	organized	74.128	selecting	104.418
used	situations	175.379	devices	55.518
	different	118.013	commonly	46.559
	learning	101.585	in	37.498
should	be	157.059	be	145.706
	she	53.101	always	38.284
	believed	51.526	why	34.073
statement	thesis	281.407	thesis	157.084
	statement	134.38	if	153.736
	if	88.637	TRUE	96.348
listening	listening	652.363	yawns	24.73
	marginal	192.869	knowledge	22.163
	appreciative	149.291	skills	20.665
help	you	103.569	you	79.075
	here	99.798	understand	44.324
	mind	92.586	will	37.971
covid	pandemic	149.556	treatments	61.092
	spread	51.017	vaccines	46.389
	disease	41.779	disease	37.717
claim	value	235.822	claim	125.297
	fact	234.758	value	114.727
	policy	234.496	fact	111.978

The analysis of collocate shifts between Grade 10 and Grade 11 reveals a potential evolution in the contextual usage and associated concepts for several key terms. In English, there appears to be a move from more active learning or modular structures towards analytical review. For research, the focus may shift from the foundational elements of a "paper" and "methodology" towards the dissemination and procedural "steps" involved in academic work, potentially referencing "journals." The context of information may have to evolved from a general act of providing and the broad concept of literacy to more active processes of managing and sharing information, such as "dissemination," "organizing," and "selecting." This shift clearly maps onto the transitioning goals of Senior High School, where students are no longer just gathering information for a basic report, but are expected to behave like actual researchers who manage, evaluate, and publish data.

The use of "used" potentially transitions from describing varied learning "situations" and "different" contexts to a greater emphasis on "devices" and the frequency of their use. For "should," the shift from discussing states of being or beliefs ("be," "she," "believed") to considering reasons, temporal aspects, or academic arguments ("always," "why," "thesis") suggests a move towards more analytical and reasoned discourse. The term "statement" maintains its association with "thesis" but shows a new emphasis on verification ("TRUE") in Grade 11. Linguistically, this marks a critical evolutionary step in student writing: language moves away from subjective beliefs ("she believed") and toward objective, reasoned academic argumentation ("thesis," "why"). Grade 11 materials are actively training students to defend claims rather than just state opinions.

A notable shift occurs with "listening," where the Grade 10 collocates suggest a focus on the act itself and related qualities, while the Grade 11 collocates, albeit with lower likelihood, hint at potential disengagement ("yawns") or learning outcomes ("knowledge," "skills"). The context of "help" seems to move from direct interpersonal assistance to a focus on comprehension, future actions, or even potentially medical interventions. Regarding "covid," the discussion appears to evolve from the initial impact ("pandemic," "spread") to responses and information sources like "treatments," "vaccines," and "articles." Finally, "claim" maintains its core associations with argumentation ("value," "fact") but sees a reduced prominence of "policy" as a direct collocate in Grade 11. These contextual shifts reflect the real-world timeline of the curriculum's creation and application. For instance, the evolution of "COVID" collocates from “pandemic” to “vaccines” and “articles” shows a transition from immediate, emotional crisis-response language to secondary, academic analysis of the event.

In general, these shifts suggest a potential trend from more foundational, descriptive, or immediate contexts in Grade 10 towards more analytical, process-oriented, and potentially abstract or specialized discussions in Grade 11. The emergence of new collocates and the changing likelihood of existing ones provide valuable insights into the evolving linguistic landscape and conceptual focus across these two educational levels.

Complementing the frequency and dispersion data, these collocate shifts offer definitive evidence of the qualitative maturation expected in a spiraling K-12 curriculum. They prove that vocabulary progression is not just about learning harder words, but about using the same words in much more sophisticated ways. As students transition from the final year of junior high school into senior high school, the linguistic landscape shifts from the descriptive to the analytical. Grade 10 modules focus on the 'what' (foundational definitions and immediate contexts), while Grade 11 modules demand the 'how' and 'why' (processes, verification, and abstract discourse). Ultimately, these collocate patterns reveal that the curriculum successfully transitions students from language consumers into critical, academic thinkers.

CONCLUSION

The primary aim of this study was to compare the lexical profiles between selected Grade 10 and Grade 11 English learning modules used in the Philippine K-12 system and determine if the Grade 11 materials demonstrated increased lexical complexity as expected within a spiraling curriculum.

The analysis revealed several key findings. Firstly, the Grade 11 corpus was substantially larger in terms of total words (tokens) and unique words (types) compared to the Grade 10 corpus, aligning with the expectation of increased content at a higher grade level. While the simple Type-Token Ratio (TTR) was slightly lower for Grade 11, suggesting potentially more repetition, more robust measures like the Moving-Average Type-Token Ratio (MATTR) indicated similar local lexical diversity. Crucially, the Measure of Textual Lexical Diversity (MTLD) scores were generally higher for Grade 11 modules, indicating greater overall lexical diversity and the ability to sustain unique vocabulary over longer text segments, thus supporting the hypothesis of increased complexity.

Keyword analysis highlighted differences in topical focus, with terms related to foundational concepts, specific tasks (like "listening," "persuasive," "argumentative"), and general academic language ("research," "information," "statement") being more prominent or 'key' in the Grade 10 corpus based on likelihood scores. Dispersion analysis (Julliand's D) showed variation in how terms were distributed, with some being spread evenly (e.g., "used," "help") while others were clustered in specific sections (e.g., "listening," "definition" in Grade 10). Collocate analysis further suggested an evolution in the contextual usage of key terms, potentially

moving from foundational or descriptive contexts in Grade 10 towards more analytical, process-oriented, or specialized applications in Grade 11 (e.g., shifts in collocates for "research," "information," "should," "covid").

In conclusion, the quantitative analysis provides empirical evidence supporting a general trend of vocabulary progression and increased lexical complexity from the Grade 10 to the Grade 11 English modules examined. While Grade 11 materials are larger and lexically richer overall, the specific keyword, dispersion, and collocate analyses reveal nuanced shifts in focus and contextual language use. These findings suggest that while the modules generally align with the expectation of increasing complexity in the K-12 curriculum, the specific nature of this progression involves not just introducing more words, but also shifts in topic emphasis and the contextual application of language.

Ultimately, these findings validate the structural design of the Philippine spiraling curriculum during a critical academic milestone. By demonstrating that vocabulary shifts from localized, descriptive units in Grade 10 to highly dispersed, analytical applications in Grade 11, this study proves that the transition from junior to senior high school successfully demands higher-order cognitive and linguistic skills.

For curriculum developers, these insights emphasize that textbook design must look beyond simple word counts to ensure that language is scaffolded contextually. For educators, understanding these collocate and dispersion trends allows for more targeted instructional strategies, ensuring students are explicitly taught how to repurpose familiar vocabulary for the specialized academic demands of senior high school and higher education.

RECOMMENDATIONS

Based on the findings of this corpus research, several key recommendations can be made to enhance the teaching and learning experience within the DepEd curriculum. Firstly, educators should be keenly aware of the increasing linguistic demands placed on students as they transition from Grade 10 to Grade 11. This awareness should inform instructional practices, encouraging the explicit teaching of more complex vocabulary and the strategic scaffolding of learning materials to support comprehension.

Curriculum developers, in turn, should consider a deliberate and gradual progression of linguistic complexity across grade levels, ensuring that the introduction of more advanced vocabulary and concepts in Grade 11 builds logically upon the foundational language skills established in Grade 10. Furthermore, the observed shift in topical emphasis suggests a need for educators to explicitly bridge the connections between foundational skills learned in Grade 10 and their application within the more specialized content of Grade 11.

A clear understanding of vocabulary distribution highlights the importance of ensuring that key academic terms are not only introduced but also consistently reinforced and applied across different contexts within the curriculum. The evolving contextual usage of key terms, revealed through the collocate analysis, underscores the need for instructional strategies that promote a deeper understanding of how language functions in different academic contexts, fostering students' ability to navigate more abstract, analytical, and specialized discourse as they progress.

Consequently, these recommendations emphasize the need for a coordinated effort between educators and curriculum developers to strategically address the increasing linguistic demands and evolving content to optimize student learning and academic success.

Finally, to build upon the empirical foundation established by this study, it is highly recommended that future researchers expand this corpus-based analysis. Subsequent studies could broaden the scope by examining learning modules across the entire secondary spectrum, from Grade 7 to Grade 12, to trace the complete trajectory of the spiraling English curriculum.

Additionally, incorporating modules from different geographical regions across the Philippines would provide a more comprehensive, nationally representative picture of vocabulary progression and linguistic complexity in the public school system.

REFERENCES

1. Ambayon, C. M. (2020). Modular-Based Approach and Students' Achievement in Literature. International Journal of Education and Literacy Studies, 8(3), 32. https://doi.org/10.7575/aiac.ijels.v.8n.3p.32
2. Bustillo, E., & Aguilos, M. (2022). The Challenges of Modular Learning in the Wake of COVID-19: A Digital Divide in the Philippine Countryside Revealed. Education Sciences, 12(7), Article 7. https://doi.org/10.3390/educsci12070449
3. Cameron, L. (2001). Teaching Languages to Young Learners. Cambridge University Press.
4. Costales, R. (2022). Productive vocabulary knowledge of pre-service teachers as a predictor of writing performance: Basis for English syllabi And instructional modules improvement. Central European Management Journal, 30(4). https://doi.org/10.57030/23364890.cemj.30.4.75
5. Department of Education. (2024). Implementing Guidelines on the School Calendar and Activities for the School Year 2024-2025. Department of Education. https://www.deped.gov.ph/wp-content/uploads/DO_s2024_009.pdf
6. Enhanced Basic Education Act of 2013. (2013). Republic Act No. 10533. The Republic of the Philippines. https://www.officialgazette.gov.ph/2013/05/15/ republic-act-no-10533/
7. Gumangan, Ma. A., & Dita, S. (2022). Multi-Word Verbs in Senior High School Academic Papers: A Corpus-Based Study. In S. Dita, A. Trillanes, & R. I. Lucas (Eds.), Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation (pp. 664–673). Association for Computational Linguistics. https://aclanthology.org/2022.paclic-1.73/
8. Hsu, W. (2009). College English Textbooks for General Purposes: A Corpus-based Analysis of Lexical Coverage. Electronic Journal of Foreign Language Teaching, 6(1), 42–62.
9. Norberg, C., & Nordlund, M. (2018). A Corpus-based Study of Lexis in L2 English Textbooks. Journal of Language Teaching and Research, 9(3), 463. https://doi.org/10.17507/jltr.0903.03
10. O'Neill, G., Moore, S., & McMullin, B. (2005). Emerging Issues in the Practice of University Learning and Teaching. 183.
11. Schools Division Office of Valenzuela. (n.d.). *Learning Modules* [under Contextualized LR Materials]. DepEd Valenzuela LRMS. Retrieved [April 12, 2025], from https://sites.google.com/deped.gov.ph/depedvalenzuelalrms/learning-resources/contextualized-lr-materials/learning-modules
12. Söğüt, M., & Vural, E. (2023). Lexical Analysis of Two ELT Coursebooks: Corpus-informed & Conventional Coursebook. Rahva Journal of Technical and Social Studies, 3(1), 32–44.
13. Yu, M., & Renandya, W. A. (2021). A corpus-based study of the vocabulary profile of high school English textbooks in China. LEARN Journal: Language Education and Acquisition Research Network. https://hdl.handle.net/10497/22789