Vocabulary Requirement Comparison Between the Verbal Reasoning Sections of the GMAT and GRE Exams

Graduate Management Admission Council

Table of contents

Executive Summary

The study used the BNC-COCA word family list to investigate the vocabulary requirements for the Verbal Reasoning sections of the GMAT and GRE. The results show that the GMAT Verbal Reasoning section required candidates to know fewer word families than the GRE Verbal Reasoning section on average. Specifically, the GMAT required 6,000 to 7,000 word families on average, whereas the GRE required 9,000 word families on average. The advantage of the GMAT over the GRE was not much regarding vocabulary requirements for reading passages. However, for non-passage items, the GMAT required 6,000 word families on average, whereas the GRE required 11,000 word families on average. The findings of the study suggest that the GMAT exam may be easier to prepare than the GRE exam for candidates who do not have a large vocabulary.

Introduction

Vocabulary is the foundation of text comprehension for any language. The proportion of words within a text known to a reader, often referred to as coverage, can directly impact the levels of comprehension. Studies have found that 98% coverage is needed for comprehending a text (Hu & Nation, 2000; Schmitt et al., 2011), and the chance of fully understanding a text with lower than 98% coverage (e.g., 95%) can significantly diminish. Therefore, for anyone to comprehend a text, they should strive to understand 98% or above of the words in the text.

Many studies have investigated the vocabulary requirements for achieving 98% coverage of the text in the English language assessments. One study found that to attain at least 98% coverage, the IELTS listening and reading sections need knowledge of 4,000 and 8,000 word families, respectively (Kaneko, 2020). The word family represents the different variations of the same word (e.g., plural and adjectives). The TOEFL listening and reading sections need the test takers to know 5,000 and 7,000 word families, respectively (Kaneko, 2020).

Previous studies have only invested in the vocabulary requirement for language assessments, and no studies have examined the vocabulary requirements for 98% coverage of the Verbal Reasoning sections of the GMAT and GRE exams, which assess the candidates' graduate-level education readiness. Although the two exams are not designed to measure candidates' language skills directly, they still measure reading comprehension and reasoning skills in English. Thus, knowing the required vocabulary levels can help non-native speakers choose suitable exams for themselves and better prepare for the exams. This study will investigate the vocabulary requirement for 98% coverage of the Verbal Reasoning sections of the two exams.

Method

438 GRE Verbal Reasoning items extracted from the most recent GRE Official Guide (ETS, 2017), Official POWERPREP Practice Exams 1 and 2 (ETS, 2024a; ETS, 2024b), and Official POWERPREP Plus Practice Exam 1 (ETS, 2024c) were included in the study. The GRE Verbal Reasoning section includes passage (i.e., reading comprehension) and non-passage (i.e., sentence equivalence and text completion) items. The GMAT Verbal Reasoning section also contains passage and non-passage items (i.e., critical reasoning items). This study randomly selected the same number of non-passage and passages items from the GMAT Verbal Reasoning item bank as those from the GRE official learning and practice materials. Note that the GMAT items are directly used in the actual exams.

We first analyzed overall coverage across all text in the GMAT and GRE Verbal Reasoning sections. Additionally, we analyzed the coverage at the item level. However, not all text at the item level was analyzed. For passage items, the words of the items often came directly from the passages that were analyzed, and their stems and options were much fewer than in the passage, making the coverage fluctuate drastically even for items associated with the same passage. Therefore, only passages were analyzed. On the other hand, stems and options of non-passage items were included in the analysis because they are integral to understanding the whole item.

Previous studies (Kaneko, 2020; Phung & Ha, 2022) used the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) word family list to investigate the vocabulary requirement. The BNC-COCA word family list was developed by Nation (2012), which includes 25,000 word families separated into 25 of 1,000 word families that represent different frequency levels of the word families in the corpora. From the first 1,000 to the 25th 1,000 word families, the most frequent to the least frequent word families in the corpus are listed. The cumulative word families can be used to represent the vocabulary levels and analyze the coverage of the text. The vocabulary level in this study was defined as the cumulative number of word families that a person needs to know. For example, a vocabulary level of 4,000 means that the person has to know the first to the fourth 1,000 word families from the word family list.

As in previous studies (Kaneko, 2020; Phung & Ha, 2022), proper nouns (e.g., names and places) were treated as known to candidates because they do not impact comprehension (Webb, 2010). Transparent compound words (e.g., blackboard) were treated as known words as well because they derive from word families that are frequently used.

Results

Table 1 shows the coverage given different vocabulary levels using the text of all items from the Verbal Reasoning sections of the two exams. It was found that to achieve 98% coverage, the GMAT Verbal Reasoning section required candidates to know 6,000 to 7,000 word families, whereas the GRE Verbal Reasoning section required candidates to know 9,000 word families.

Table 1: Overall Coverage by Vocabulary Level for the GMAT and GRE Verbal Reasoning Sections

Level	GMAT	GRE
1k	69.9%	67.3%
2k	84%	80.5%
3k	93.6%	90.8%
4k	96%	93.6%
5k	97.1%	94.9%
6k	97.9%	96.1%
7k	98.4%	97%
8k	98.7%	97.7%
9k	98.9%	98%
10k	99.1%	98.4%
11k	99.3%	98.7%
12k	99.4%	99%
13k	99.5%	99.2%
14k	99.6%	99.2%
15k	99.6%	99.4%
16k	99.7%	99.4%
17k	99.7%	99.5%
18k	99.7%	99.5%
19k	99.8%	99.5%
20k-25k	99.8%	99.6%

Tables 2 and 3 summarize the item-level coverage by vocabulary level for the GMAT and GRE exams, respectively. The results showed that for the GMAT exam, 98% coverage required the knowledge of 6,000 to 7,000 word families on average. For the GRE exam, 98% coverage required the knowledge of 11,000 word families on average.

Table 2: Descriptive Statistics of Coverage by Vocabulary Level for the GMAT Verbal Reasoning Section

Level	Mean	Median	SD	Min	Max
1k	71%	71.1%	6.4%	49.2%	92.1%
2k	85%	85.2%	5.5%	62.2%	100%
3k	93.7%	94.6%	4.5%	80.2%	100%
4k	96.1%	97.2%	3.7%	82.7%	100%
5k	97.2%	98.2%	3%	86.3%	100%
6k	97.9%	99%	2.6%	86.3%	100%
7k	98.4%	99.4%	2.3%	88.8%	100%
8k	98.7%	99.5%	2%	89.6%	100%
9k	98.9%	100%	1.9%	90.6%	100%
10k	99.2%	100%	1.6%	90.6%	100%
11k	99.3%	100%	1.5%	90.6%	100%
12k	99.4%	100%	1.4%	90.6%	100%
13k	99.5%	100%	1.2%	91.2%	100%
14k	99.6%	100%	1.1%	92.3%	100%
15k	99.6%	100%	1%	92.3%	100%
16k	99.7%	100%	0.9%	92.3%	100%
17k	99.7%	100%	0.9%	92.3%	100%
18k	99.7%	100%	0.9%	92.3%	100%
19k	99.7%	100%	0.9%	92.3%	100%
20k	99.7%	100%	0.8%	92.3%	100%
21k-25k	99.8%	100%	0.8%	92.3%	100%

Table 3: Descriptive Statistics of Coverage by Vocabulary Level for the GRE Verbal Reasoning Section

Level	Mean	Median	SD	Min	Max
1k	63.7%	64.4%	8.8%	38.1%	83.3%
2k	76.2%	77.6%	8.2%	45.5%	94.7%
3k	86.4%	87.5%	7.2%	54.5%	100%
4k	89.9%	90.9%	6.4%	59.1%	100%
5k	91.8%	92.5%	6%	59.1%	100%
6k	93.3%	94.4%	5.5%	71.4%	100%
7k	94.9%	96.1%	4.9%	72.7%	100%
8k	96.1%	97.3%	4.4%	72.7%	100%
9k	96.7%	97.6%	3.9%	77.3%	100%
10k	97.4%	98.3%	3.4%	81.8%	100%
11k	98%	99%	2.8%	82.6%	100%
12k	98.5%	100%	2.4%	82.6%	100%
13k	98.8%	100%	2.1%	82.6%	100%
14k	99%	100%	1.8%	90.5%	100%
15k	99.1%	100%	1.6%	90.5%	100%
16k	99.2%	100%	1.4%	93.5%	100%
17k	99.3%	100%	1.3%	93.5%	100%
18k	99.4%	100%	1.2%	93.5%	100%
19k	99.4%	100%	1.2%	93.5%	100%
20k	99.5%	100%	1.2%	93.5%	100%
21k	99.5%	100%	1.2%	93.5%	100%
22k-25k	99.5%	100%	1.1%	93.5%	100%

Figures 1 and 2 show the item-level coverage by vocabulary level for passages and non-passage items of the GRE and GMAT exams. The coverage of passages of different vocabulary levels was similar between the two exams. However, the coverage of the non-passage items was different. Specifically, the coverage of the non-passage items in the GRE showed much higher variability than that of the non-passage items in the GMAT exam. Table 4 summarizes the mean coverage by vocabulary level, item type, and exam type. The differences in vocabulary requirements for 98% coverage of reading passages between the GMAT and GRE were small. On average, the GMAT required knowledge of 7,000 word families, whereas the GRE required knowledge of 8,000 word families. However, the requirements for 98% coverage were drastically different for non-passage items of the two exams. Specifically, the GMAT non-passage items required knowledge of 6,000 word families, and the GRE non-passage items required knowledge of 11,000 word families on average.

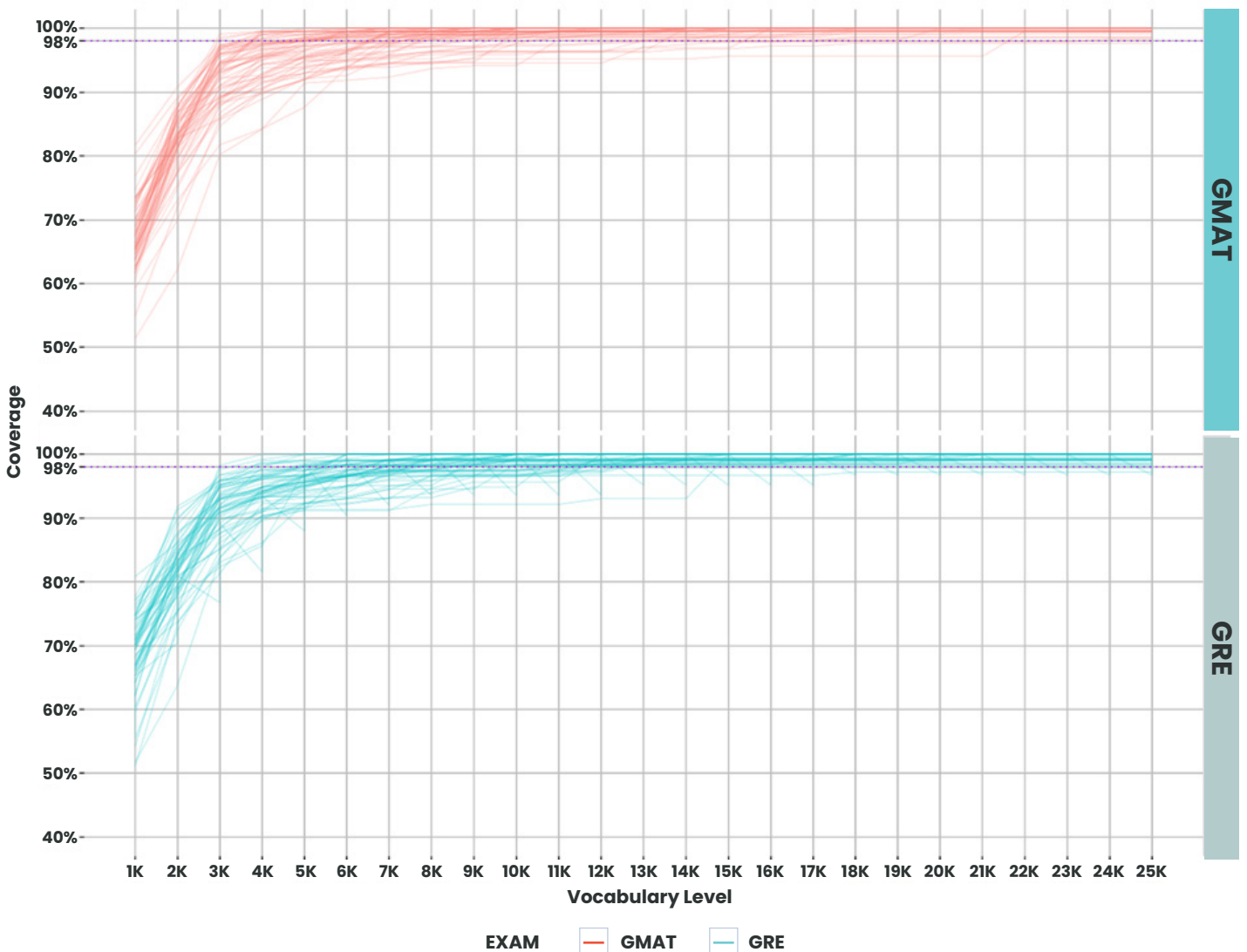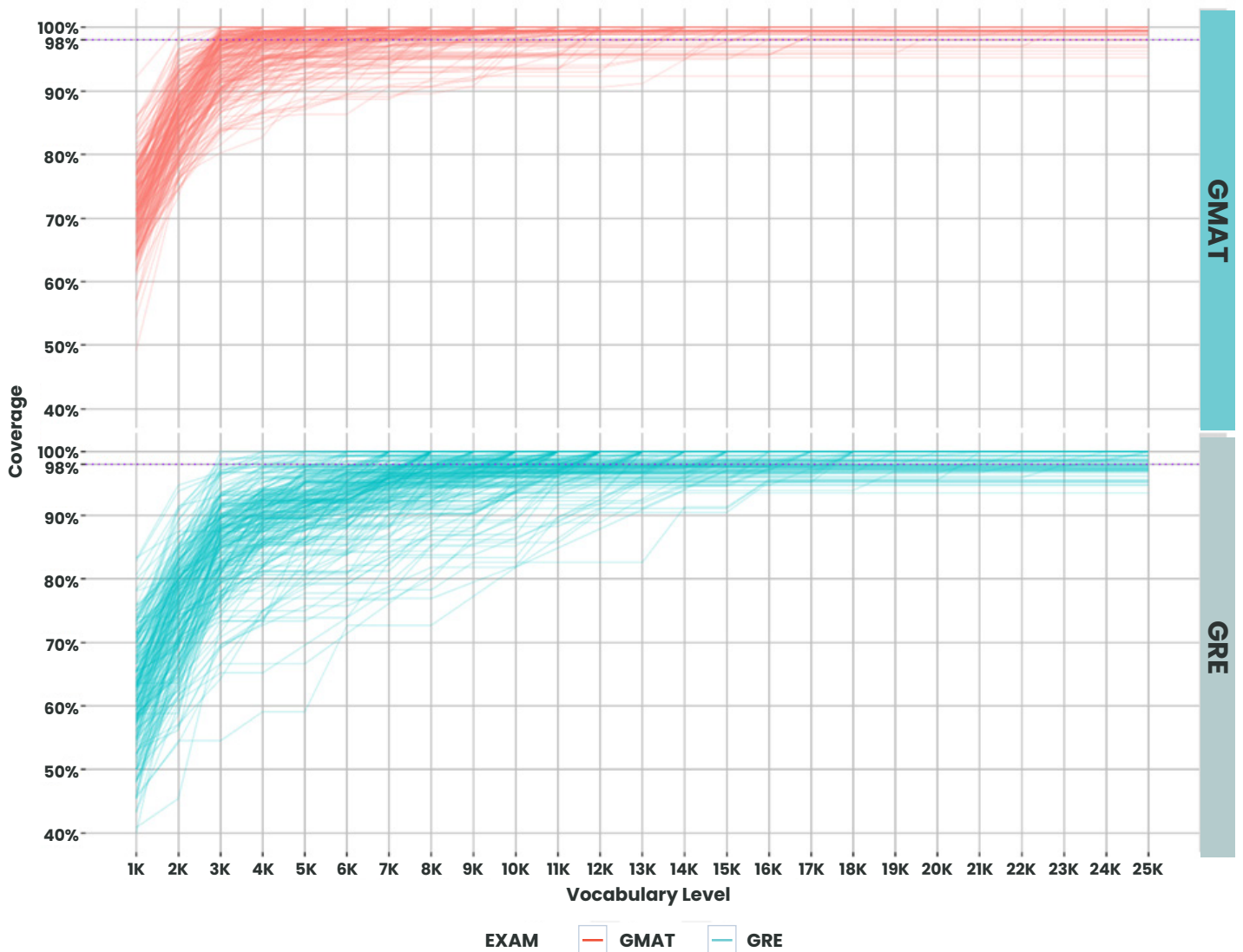## Figure 1: Coverage of Passages by Vocabulary Level and Exam Type

Figure 2: Coverage of Non-Passage Items by Vocabulary Level and Exam Type

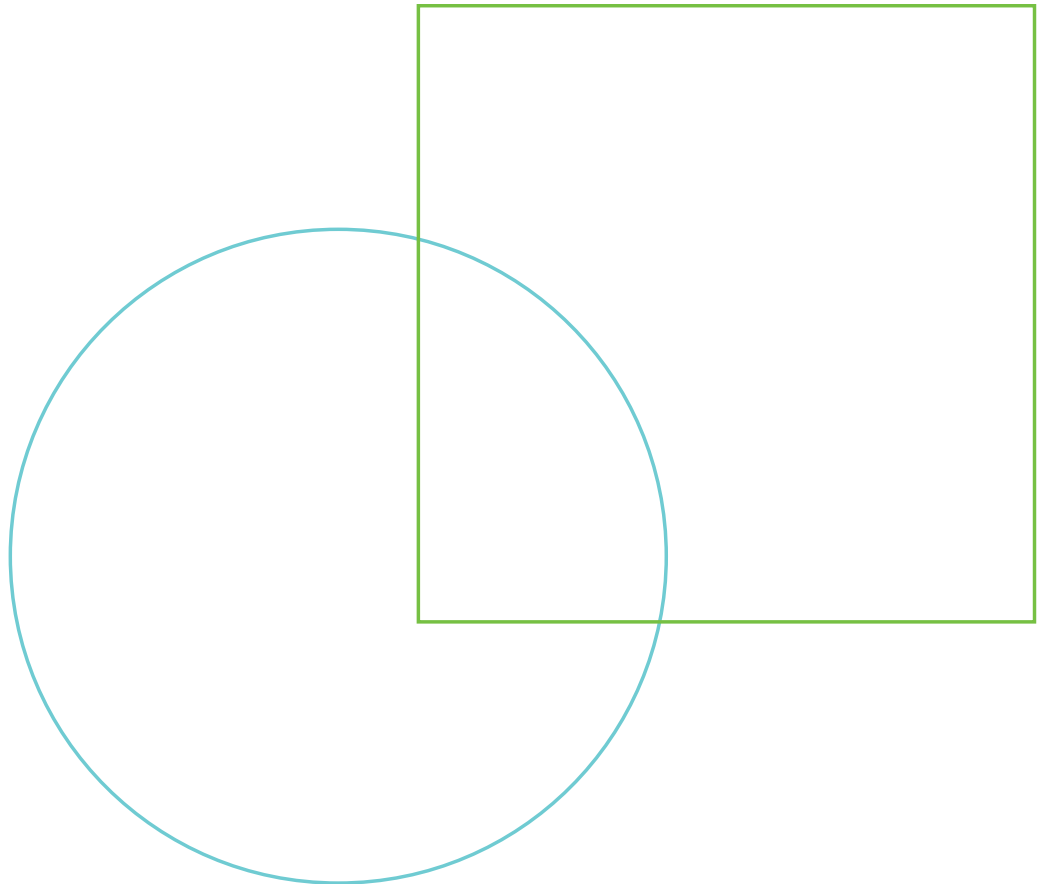## Table 4: Mean Coverage by Vocabulary Level, Item Type, and Exam Type

Level	GMAT		GRE	
	Non-Passage	Passage	Non-Passage	Passage
1k	72%	67.5%	62.3%	68.5%
2k	85.7%	82.5%	74.8%	81.4%
3k	94.1%	92.4%	85.1%	90.9%
4k	96.3%	95.1%	88.7%	93.9%
5k	97.4%	96.4%	90.7%	95.6%
6k	98%	97.4%	92.3%	96.7%
7k	98.5%	98.1%	94.1%	97.4%
8k	98.8%	98.5%	95.5%	98%
9k	99%	98.7%	96.3%	98.3%
10k	99.2%	99.1%	97.1%	98.6%
11k	99.3%	99.2%	97.8%	98.7%
12k	99.5%	99.3%	98.3%	99%
13k	99.5%	99.4%	98.7%	99.2%
14k	99.6%	99.5%	98.9%	99.2%
15k	99.7%	99.5%	99.1%	99.4%
16k	99.7%	99.6%	99.2%	99.4%
17k	99.7%	99.6%	99.3%	99.4%
18k	99.7%	99.7%	99.4%	99.5%
19k	99.7%	99.7%	99.4%	99.5%
20k	99.8%	99.7%	99.4%	99.5%
21k-25k	99.8%	99.7%	99.5%	99.6%

Summary

The overall results demonstrated that the GMAT Verbal Reasoning section required 6,000 to 7,000 word families to reach 98% coverage of the text compared to the 9,000 words required for the GRE Verbal Reasoning section. The significant differences stemmed from non-passage item types. Specifically, for non-passage items, the GRE required 11,000 word families, whereas the GMAT only required 6,000, 5,000 word families fewer than the GRE. However, the difference between the GMAT passage vocabulary requirement (i.e., 7,000) and the GRE passage vocabulary requirement (i.e., 8,000) was small. The results indicate that GRE test takers need a more extensive vocabulary than the GMAT test takers. The GMAT may be easier to prepare for than the GRE for candidates who do not have a large vocabulary (e.g., English learners).

A previous study discovered that the IELTS and TOEFL reading comprehension items require knowledge of 8,000 and 7,000 word families, respectively (Kaneko, 2020). In conjunction with this study's findings, we can conclude that candidates who did well on the reading comprehension parts of the IELTS and TOEFL would have no problem preparing for the GMAT Verbal Reasoning section. However, they may need to increase their vocabulary to do well in the GRE Verbal Reasoning section.

Contact Information

If you have questions or comments about the study findings, methodology, or data, please contact Yanyan Fu, a Psychometrician, at **yfu@gmac.com**, or Kyung (Chris) T. Han, Senior Director and head of Test Development and psychometrics, at **khan@gmac.com**.

References

ETS. (2017). The Official Guide to the GRE® General Test (3rd Edition). McGraw Hill.

ETS. (2024a). POWERPREP® Online - Practice Test 1. Retrieved from
https://ereg.ets.org/ereg/testPrep/viewtestPreparation?_p=GRI

ETS. (2024b). POWERPREP® Online - Practice Test 2. Retrieved from
https://ereg.ets.org/ereg/testPrep/viewtestPreparation?_p=GRI

ETS. (2024c). POWERPREP PLUS® Online - Practice Test 1. Retrieved from
https://ereg.ets.org/ereg/testPrep/viewtestPreparation?_p=GRI

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. Reading in a Foreign Language, 13(1), 403–430.

Kaneko, M. (2020). Lexical frequency profiling of high-stakes English tests: text coverage of Cambridge First, EIKEN, GTEC, IELTS, TEAP, TOEFL, and TOEIC. JACET Journal, 64, 79-93.

Nation, I.S.P. (2012). The BNC/COCA word family lists. Retrieved from
https://www.eapfoundation.com/vocab/general/bnccoca/

Phung, D. H., & Ha, H. T. (2022). Vocabulary demands of the IELTS Listening Test: An in-depth analysis. Sage Open, 12(1).
https://doi.org/10.1177/21582440221079934

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. The Modern Language Journal, 95(1), 26–43.

Webb, S. (2010). Using glossaries to increase the lexical coverage of television programs. Reading in a Foreign Language, 22, 201–221.

Authored by Yan Fu & Chris Han

Graduate
Management
Admission
Council™