



A frequency-based wordlist of Japanese junior high school textbook vocabulary

Ryan Klinger

Temple University, Philadelphia, USA
ryandklinger@gmail.com

Abstract

Several studies have examined the content of Japanese junior high school textbooks in relation to various frequency-based word lists such as the New General Service List (NGSL) (Browne et al., 2013) (e.g, Nakayama, 2022a,b) and the British National Corpus (BNC) (e.g, Wongsarnpigoon, 2018), and have identified potential issues in terms of lexical coverage and the repetition of vocabulary items. Considering the potential differences in individual textbook series and differences between frequency-based wordlists, it is worthwhile to evaluate the entirety of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) approved junior high school textbooks (MEXT, 2024) in relation to multiple researched word lists including the NGSL, and the New JACET8000 (JACET, 2016) wordlist. To do so, a corpus of 18 textbooks from 6 publishers was analyzed using the software AntConc (Anthony, 2022) to generate a junior high school vocabulary wordlist sorted by the combined frequency and range of each item. While MEXT has advanced the target number of vocabulary items junior high school (JHS) students in Japan are expected to know entering high school to a range between 2,200 and 2,500 words (MEXT, 2019), the content of those words is relatively unknown. Thus, the aim of this study is to evaluate the content of the collective vocabulary items within the catalogue of MEXT approved

JHS textbooks, by examining their range, frequency, and difference in ranking between the NGSL and JACET8000 wordlists. Additionally, the pedagogical applications derived from the analysis of a JHS textbook wordlist are discussed.

Keywords: JACET8000 Wordlist, Japanese EFL Textbooks, Junior High School, New General Service List, Textbook Corpus, Vocabulary

Introduction

Standardized textbooks are one source of EFL vocabulary that reflects the vision and curriculum goals of a larger educational system. In Japan, there are 18 English language textbooks from six different publishers that have been approved by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for junior high school (JHS), per the catalogue of textbooks (MEXT, 2024). As these textbooks have been approved following the reforms to English curriculum standards, which saw an increase in the target number of vocabulary items a graduating JHS student is expected to acquire from 1200 items to a range between 1600 to 1800 words (MEXT, 2019), it is worthwhile to examine the content of these words, both internally through measurements of range and frequency, as well as externally through comparisons to corpus-based wordlists such as the New General Service List (NGSL) (Browne et al., 2013) and the JACET8000 wordlist (JACET, 2016).

Through understanding the frequencies and ranges of EFL textbook vocabulary for Japanese JHS students, it is possible to gain insights to the universality of the language used, as well as the number of items that are exclusive to a particular coursebook series. Essentially there are six individual routes a JHS student in Japan takes before entering high school, potentially with different levels of variance in the vocabulary exposed to through textbooks between each. While vocabulary acquisition is not exclusively restricted to the content of textbooks, an analysis of the repetition of words and their availability offers insights to the degree in which acquisition is supported and informs instructors on how best to utilize textbook vocabulary lists for learning based on the results of those examinations.

There are also several benefits in comparing the vocabulary content of Japanese JHS English textbooks to corpus-based wordlists, such as the NGSL and JACET8000. As researched wordlists are based on a large corpus ordered by word frequency, such lists serve as measuring sticks for lexical coverage that provide the thresholds for reading comprehension and reading fluency (Schmitt & Schmitt, 2014; Vilkaitė-Lozdienė & Schmitt, 2019). Thus, the degree of alignment between wordlists illuminates the relative value of the items JHS students in Japan are being exposed to. Furthermore, profiling tools that allow instructors to evaluate EFL materials are also based on researched lists, such as the NGSL Profiler Tool (Browne, 2024), and the New Word Level Checker (NWLC) (Mizumoto, 2021), which are useful if the learner is expected to have adequate knowledge of words within the

bands of those wordlists. A comparison between a compiled JHS textbook list and NGSL and JACET provides insights into the congruency of such lists, as well as any potential gaps.

Literature Review

Foremost in regard to the internal analysis of vocabulary within a text, there is plenty of research that covers the benefits of frequency or word repetition as well as the benefit of learning high frequency words. Webb (2007) found that sizable gains in vocabulary knowledge occur when a learner encounters an item ten times within context and suggests that more repetitions is beneficial for acquisition. There are, of course, numerous variables that affect the relationship between repetition and vocabulary learning. While research on word repetition is more abundant for incidental learning, as is the case with Webb (2007), explicit vocabulary learning has shown to require less repetitions, from as low as four (Teng & Xu, 2022; Zhang, 2022). Waring and Takaki (2003) found that incidental vocabulary learning from a graded reader required different encounter rates depending on the type of word knowledge analyzed, from as low as eight encounters for form recognition, to as high as 18 or more for meaning recognition, three months after exposure. This variance is also reflected in a meta-analysis conducted by Uchiyama et al., (2019), which suggests that up to 20 repetitions might be required for effective vocabulary acquisition, depending on age, learner vocabulary knowledge, spacing, and visual support among others.

Ultimately, within the context of JHS English textbooks in Japan, the repetition of items has no concrete benchmark. After all, the explicit instruction of vocabulary items, while perhaps influenced by textbook word banks and word lists, remains at the discretion of the instructor, as well as the reading mode, whether students are reading intensively or extensively. Thus, the optimistic benchmark for repetitions considered for the analysis in this study is aligned with the Webb (2007) findings.

The other factor for internal analysis related to the present study is the aforementioned benefit to learn high frequency words, as they provide excellent coverage of English texts (Kremmel 2016; Nation, 2022), and have also been found, in the case of the NGSL, to provide adequate coverage for university entrance exams in Japan (Iwamoto & Kramer, 2020). Additionally, a comparison between the NGSL and JACET lists, conducted by Ishikawa (2019), revealed a general congruency between the two, pertaining to the first 2000 words of each list. Therefore, for discussions regarding high frequency words, there is little variance between the researched lists.

Studies examining the vocabulary contents of Japanese JHS English textbooks are generally separated into extensive examinations of approved textbooks as a whole, or intensive studies that focus on particular aspects evidenced in a particular textbook series. One extensive study that included 28 textbooks across elementary school and junior high school in Japanese was conducted by Mikajiri (2023). Their research aimed to compare the lexical bundles within the textbooks to an English spoken corpus (SUBTLEXus). One finding that emerged from the study was that there were inconsistencies between the spoken English corpus and lexical bundles within the textbooks, and based on such results, it was suggested that textbook writers should

consult corpora data to improve language learning materials, to better align actual language use and the language learned in textbooks.

Another extensive examination of Japanese JHS English textbooks was done by Muraoka (2010), who observed the frequencies and ranges of words within the MEXT approved textbooks at the time, of which there were 18 total textbooks from 6 different publishers. The analysis showed that nearly 50% of the vocabulary items had a range of 1 (meaning they were present in only one textbook series), and only 20% were found in every textbook series. For frequency, Muraoka found that over 600 words in the second and third grade textbooks were repeated once, and across all textbooks, 51.7% of all words had a frequency value between 1 and 2. However, the caveat of this study is that it did not utilize the word family or lemma word counting unit, which groups words based on their derivational forms or part of speech respectively. Therefore, the frequency data might be less severe than reported, as morphological forms of the words were counted as separate items (e.g., eat and ate were considered separate items).

For intensive textbook analysis, Nakayama (2022b) examined a particular series of Japanese JHS English textbooks, *Here WE Go!* (Oda et al, 2021a,b,c), and compared it to the NGSL (Browne et al., 2013). While the NGSL provided good coverage of the textbooks at 95%, the coverage the textbooks gave for the NGSL was less so, at 37%. Nakayama concluded that a student attempting extensive reading, for example, would require supplemental word knowledge to make up for the NGSL words that were not found in the *Here WE Go!* Textbooks. Similarly, Wongsarnpigoon (2018) analyzed a series of JHS textbooks, *New Crown* (Takahashi et al. 2012), a previous version of the textbook that is currently utilized in JHS. Congruent with the findings issued by Nakayama (2022b), Wongsarnpigoon found that while there was adequate representation for the first 1,000 words of the British National Corpus (BNC), only 996 of the first 2,000 words of the list were found within the textbook series. Additionally, aligning with Muraoka (2010) and Nakayama (2022b), the repetition of items was noted as a particular concern, indicating that several iterations of JHS textbooks have not necessarily advanced in terms of reusing words across units and textbooks.

Outside of the JHS textbook context, a few studies have examined elementary and high school textbooks with similar scopes in Japan. Nakayama (2022a) compared nine senior high school English textbooks from three different publishers to the NGSL and found similar results to his subsequent studies for JHS textbooks (Nakayama, 2022b): the NGSL provided adequate coverage with each analyzed text having lexical coverage above 92%, yet the coverage the textbooks provided of the NGSL was much less so, with coverage percentages ranging from 12% to 38%. Hoshino (2020), comparable to the analysis conducted by Muraoka (2010), examined 14 elementary textbooks from seven publishers for fifth and sixth grade in order to tabulate the ranges of types that appeared across each series. 275 types (12%) had a range of seven, and 969 types (42%) had a range of 1. These studies indicate that the general frequencies and ranges of vocabulary items in textbooks are not exclusive to JHS but are common across publishers in three levels of education in Japan.

In summation, there are certain trendlines that have carried over various versions of JHS English textbook implementation in Japan, such as a large quantity of low repetition items, and items that are exclusive to particular textbooks. Individual

textbook series have also shown drop-offs in terms of representing items from large corpus wordlists, particularly following after the first 1,000 words such as with the NGSL (Nakayama, 2022b), and the JACET8000 (Klinger, 2024). However, while MEXT has increased the number of target vocabulary items for JHS students from 1200 items to a range between 1600 to 1800 items (in addition to the 600 to 700 items included in the new elementary curriculum) (MEXT, 2019), it is also important to understand that textbook criteria are not exclusively aimed at word content but rather is evaluated differently between each board of education (Hammett, 2020). Even so, the purpose of this study is to provide a broader picture of the content of JHS English textbook words, with the purpose to investigate the following questions:

1. What are the frequencies and ranges of items from a collective JHS textbook wordlist?
2. How does the ranking of items on a JHS wordlist compare to the JACET8000?
3. How does the ranking of items on a JHS wordlist compare to the NGSL?

Through answering these questions, a better understanding of the vocabulary that Japanese JHS students are expected to learn is achieved, as well as providing insights towards the applicability of such data in regard to material development and adaptation, as well as efforts to synergize early English education in JHS to subsequent grade levels.

Methods

The first parameter that was established in creating the JHS textbook wordlist was to identify the textbooks that qualified for analysis. This was derived from the Junior High School Textbook Index published by MEXT (2024). From this catalogue of approved texts, the eighteen textbooks utilized in this study were determined: *New Horizon* series (Tokyo Shoseki), *Sunshine* series (Kairyudo), *New Crown* series (Senseido), *One World* series (Kyoiku-Shuppan), *Here We Go!* series (Mitsumura Toshio), and *Blue Sky* series (Shinko-Keirin). These coursebooks were then manually typed into separate word document files, with one document file per book, with a total of 18 individual documents. The primary data collected from the textbooks included written dialogues, exercise prompts, language examples for grammar, vocabulary lists, and literature. Audio files, glossaries, and appendices were not examined, and are excluded from the data presented in this study.

After typing up the documents, all 18 initial drafts of the typed up textbooks were then placed into AntConc (Anthony, 2022) for the purpose of screening the documents for spelling errors, inflectional forms, as well as derivational forms. AntConc sorted the words into types, allowing for common mistakes across documents to be identified. Following this, each individual document was placed into AntConc separately, and each word was checked, with necessary changes made to the documents using Microsoft Word's find and replace function. Once all the documents had been screened, the updated drafts were once again placed into AntConc to generate a wordlist sorted by frequency.

The words, their frequencies within the 18-textbook corpus, and their ranges across the corpus, were copied from AntConc and pasted into an Excel spreadsheet. The frequency of each word was then multiplied by its range to generate the item's word power. The words were then sorted by the word power value, and the rank of each word was determined. This spreadsheet was the initial draft of the JHS textbook wordlist.

Following that, a second screening was conducted through comparing the JHS textbook wordlist to the JACET8000 wordlist, which was selected due to its length (8000 items), as well as being intended for Japanese learners (Ishikawa, 2019). In order to diminish the effect of proper nouns on the comparisons between wordlists, all proper nouns remaining on the JHS list were removed, with the exception of “Japanese” and “English”, as these words were included on the JACET8000 list. Additionally, several Japanese origin words were also cut from the list (e.g., sushi, Tsunami), due to their ambiguous nature related to English vocabulary and being excluded from the JACET8000 list, whereas “typhoon” was retained due to having a separate orthographical form.

To compare the JACET8000 wordlist to the JHS textbook wordlist, the words copied from the AntConc generated list in the spreadsheet were searched using Excel's search feature within the Excel version of the JACET8000 list, to determine the rank of the corresponding word on the JACET8000 list. The rank was then typed into a separate column on the JHS wordlist spreadsheet. For items that were not found on the JACET list (off-listed items), the JACET rank value was set at 8001. The purpose of this value is to indicate a threshold between items included on the JACET list, and items excluded from the JACET list. This process allowed for several items on the JHS list to be adjusted to better align with the JACET list. In some cases, the context of individual words was checked using AntConc's KWIC (keyword in context) function, to determine if an entry needed to be separated into different items. Such examples of this included “train” (noun) and “training” (verb), “rose” (noun) and “rise” (verb) and “leaf” (noun) and “leave” (verb). It should be noted that JACET sometimes distinguishes verbal forms from their respective nouns (e.g., skiing and ski), but in other cases it does not (e.g., hunt and hunting). For comparative purposes, the JHS wordlist attempted to align with the JACET format as much as possible in terms of how entries were conjoined or separated. In this regard, it is difficult to pinpoint a specific word-counting unit for this study, as some entries abide by part of speech, and others do not. Aside from the aforementioned cases, however, the remainder of the list abides by the flemma counting unit.

After filling out the JACET ranks on the spreadsheet, the same process was conducted for the NGSL wordlist, which was selected due to its widespread use in research and graded texts, with off-listed words given a similar value of 8001, to make it easier to align the comparisons between both datasets. The NGSL utilized in this study was the updated 1.2 version (Browne et al. 2024), however, the difference between the updated version and previous versions was not a consideration for this study.

With the primary data collected in the Excel spreadsheet, values were graphed and charted to answer the proposed research questions. First, the word power (frequency and range), and rankings of the JHS were compiled. For the JACET and NGSL comparisons, the recorded rankings were graphed in four separate bands (of approximately

891 items per band), for the purpose of identifying shifts in the data incrementally and creating a better visual representation of these trends.

The difference values between the rankings were then tabulated, and the number of items that had a gap value of greater than 1000 were added to a separate chart to fully illustrate the degree of deviation between the JACET and NGSL from the JHS textbook wordlist across each of the four bands. The Spearman's rho correlations and related values were calculated using an online tool (langtest.jp) (Mizumoto, 2015).

Results

The final version of the JHS wordlist compiled contained 3,567 items sorted by word power, a value determined by the frequency within the analyzed textbook corpus multiplied by its range. 5% of the words on the list have a range of 18, and 34% of the items have a range of 1. In terms of frequency, 40% of the items have a frequency of 10 or greater, 61% of items have a frequency of 5 or greater, and 10% of the items have a frequency of 1.

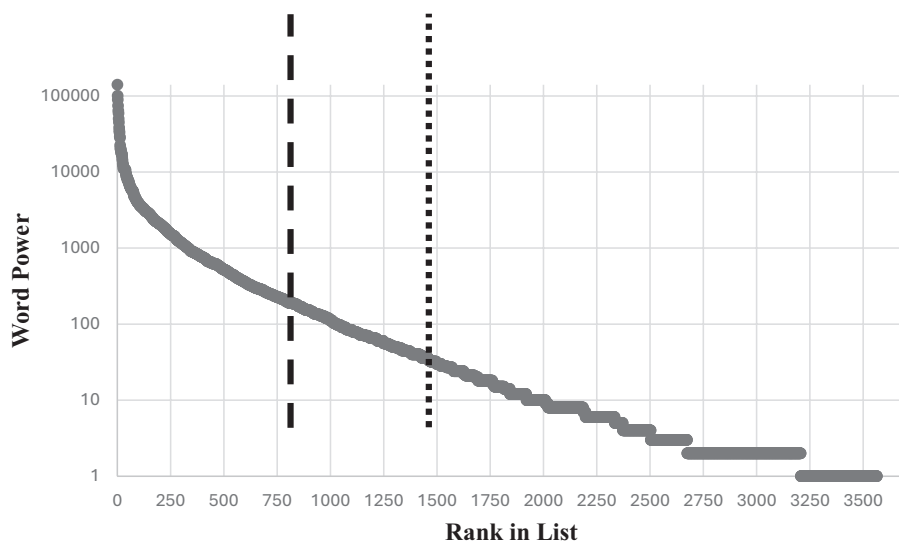


Figure 1 *Word power and ranking of the JHS Textbook Wordlist.*

A deeper examination of the range values from the JHS wordlist revealed that 802 items (22%) have a range of 6 or greater where each item appears within at least every textbook series, as illustrated by the dashed line in Figure 1. The dotted line in Figure 1 represents the threshold for items with a frequency of 10 or greater. In other words, the items to the left of the dashed range line are representative of items that all JHS students are adequately exposed to, items between both lines are adequately frequent, but not encountered in every textbook series, and the items to the right of the dotted line are both poorly distributed between textbooks, and inadequately frequent.

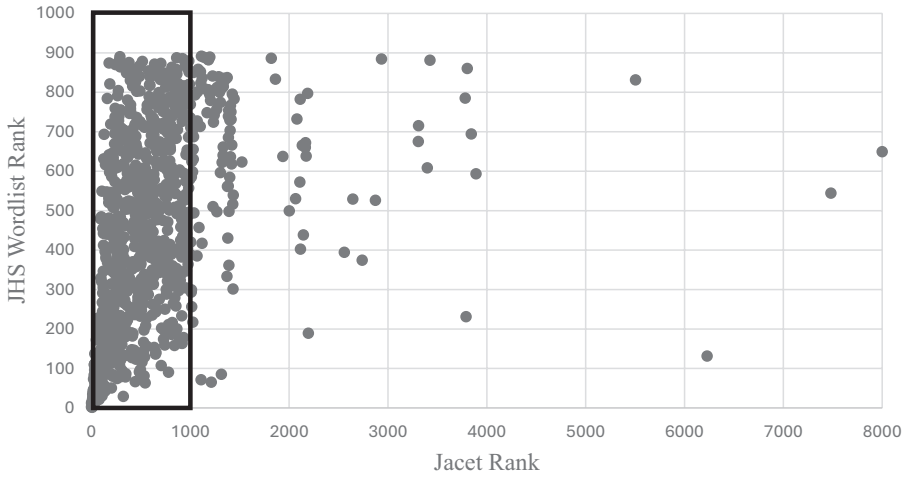


Figure 2 *Ranking comparisons between JACET8000 and JHS Wordlist's 1st band.*

The comparison between the first band of the JHS wordlist and the JACET8000 wordlist indicates a relative congruency between the rankings of items, as illustrated by Figure 2. The absolute values of differences between word rankings revealed that only 39 items (4%) were 1,000 or greater, and 143 items (16%) were 500 or greater. The ranked correlation (Spearman's rho) for the first band of the JHS wordlist compared to JACET was moderate ($p = .58$), with 95% confidence intervals (CIs) between 0.53 and .61, and a p -value = 0.

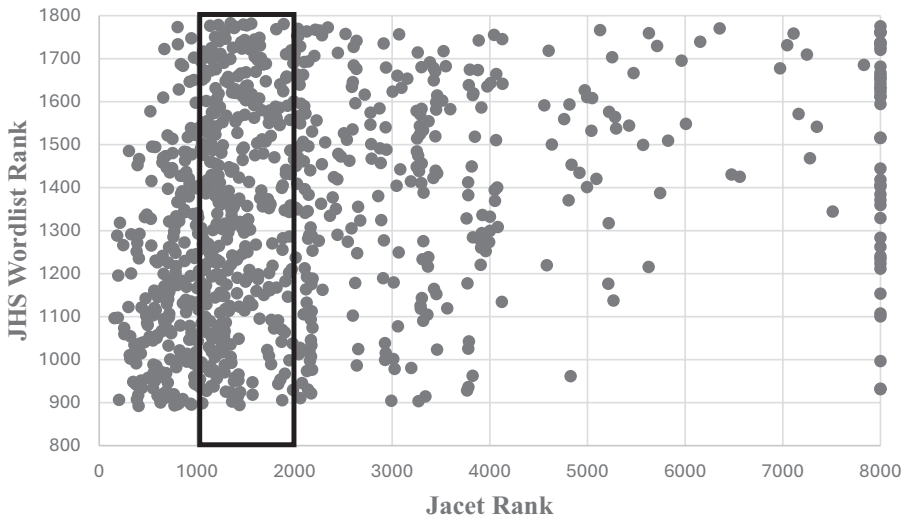


Figure 3 *Ranking comparisons between JACET8000 and JHS Wordlist's 2nd band.*

In the case of the second band of the compiled JHS wordlist, 250 items (28%) had a difference value greater than 1,000, as illustrated in Figure 3. 97 (10%) of those items had a difference value between 1000 and 2000, 65 (7%) are between 2000 and 3000, and 88 (9%) are greater than 3000. Conversely, 41% of the words within this band had a negative difference value, indicative of items that have a higher ranking on the JACET list, which is represented in Figure 3 by items that are left of the black rectangle. The ranked correlation for the second band was also significantly reduced, with a Spearman's rho value of $p = .36$ with 95% CIs between 0.3 and .41, and a p -value = 0.

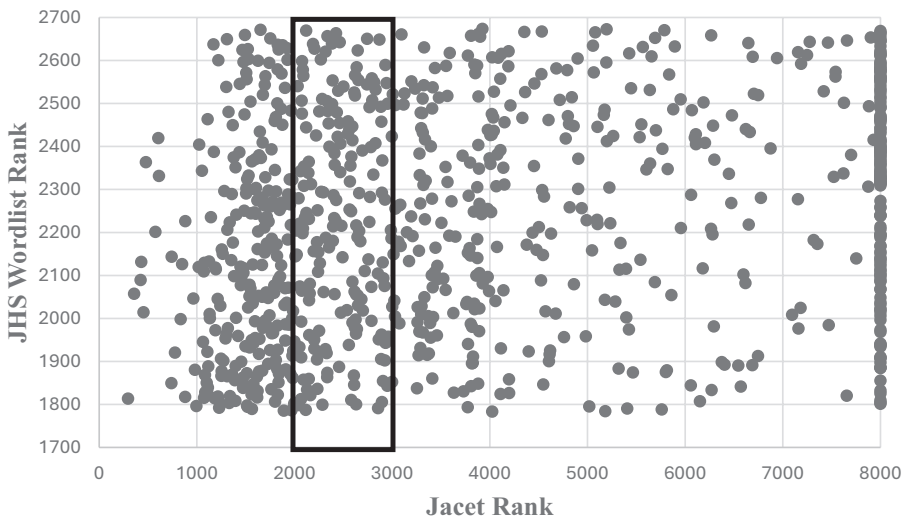


Figure 4 Ranking comparisons between JACET8000 and JHS Wordlist's 3rd band.

The comparison between the third band of the JHS wordlist and JACET illustrates further separation between the two lists terms of rank as shown by Figure 4. 48% of items within this band have a difference value of a 1000 or greater, with 29% of those words having values between 1000 and 2000, 13% between 2000 and 3000, and 56% having values greater than 3000. There is also a decrease in the number of items that have a higher JACET rank than the JHS wordlist, down to 32%, as well as a further decrease in the ranked correlation values ($p = .25$), with 95% CIs between 0.18 and .3, and a p -value = 0.

The fourth and final band of the JHS wordlist, when compared to the JACET wordlist, further continues the trend of an increasing number of items with a difference value of 1000 or greater, albeit not to the degree as seen between the second and third bands. 55% of the items have a difference value greater than 1000, with 18% of those being between 1000 and 2000, 13% between 2000 and 3000, and 68% having values greater than 3000. As shown in Figure 5, the disbursement of items is more

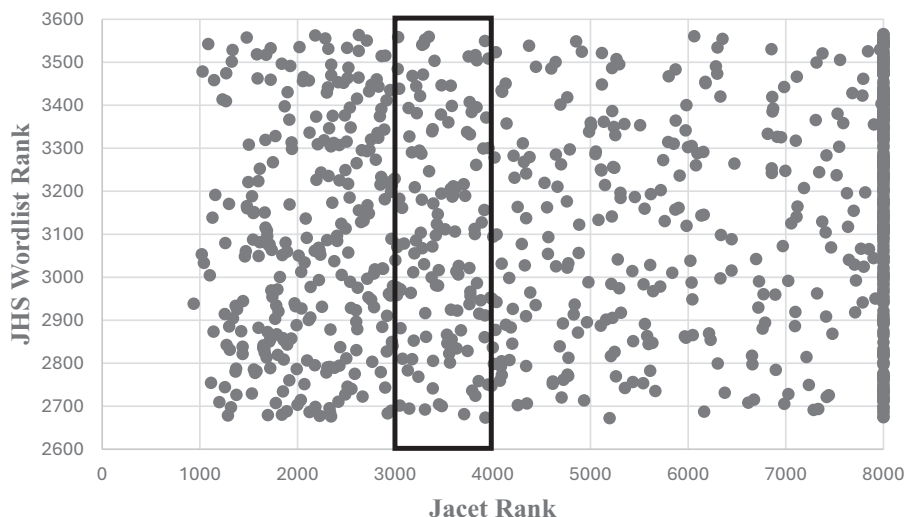


Figure 5 Ranking comparisons between JACET8000 and JHS Wordlist's 4th band.

pronounced, with the exception of off-listed words that are visually represented as a near straight line at the edge of the chart. Similar to the analysis of the 3rd band, the number of items that have higher JACET rankings remained relatively the same, at 31%, with the ranked correlation values decreasing further at ($p = .14$) with 95% CIs between .07 and .2, with a p -value = 0.

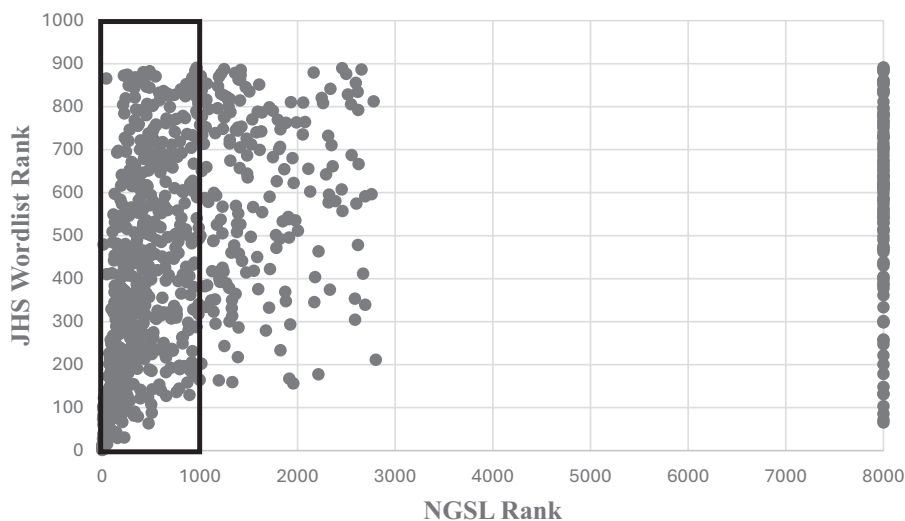


Figure 6 Ranking comparisons between the NGSL and JHS Wordlist's 1st band.

The comparison between the NGSL and the JHS wordlist nets different results within the first band, as 22% of the items have an absolute value difference greater than 1000, which is 18% more than with the JACET list. As shown in Figure 6, given the size of the NGSL list, there are expectedly more off-listed items, yet overall, the trendline suggests congruency between the JHS wordlist and the NGSL for the first band. The ranked correlation values are comparable to the first band results for the JACET list, with a Spearman's rho value of $p = .58$, CIs between .53 and .62, and a p -value = 0.

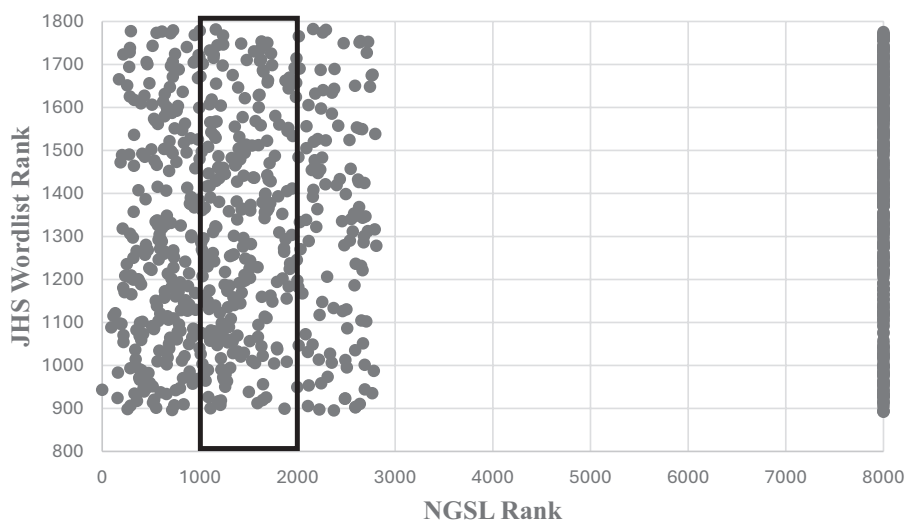


Figure 7 Ranking comparisons between the NGSL and JHS Wordlist's 2nd band.

The distribution of words in the second band of the JHS wordlist when compared to the NGSL, as shown in Figure 7, is more definite. 43% of the items in this band have difference values greater than 1000. Contrarily, 34% of the items in this band consisted of words that are more highly ranked on the NGSL than on the JHS list, as indicated by the items shown in Figure 7 left of the box representing the 2nd band of the JHS wordlist. Furthermore, the ranked correlation measurements are more prominently affected by the off-listed items, with a power value of ($p = .16$), CIs between .09 and .22, and a p -value = 0.

In the case of the third band of the compiled JHS wordlist, the number of items with a difference value greater than 1000 is 56%, a steady increase in value from the second band. The visual data represented in Figure 8 reveals a more even distribution of words, and a weakening of a noticeable trendline between the compared rankings. 28% of the items are shown to the left of the box indicating items that belong to the first 2,000 words of the NGSL, with 14% being contained in subsequent 1,000 words of the list.

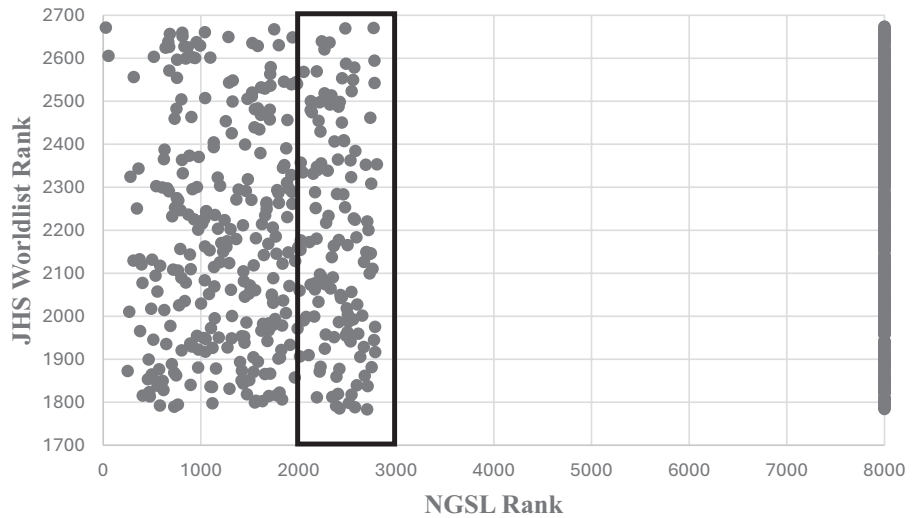


Figure 8 *Ranking comparisons between the NGSL and JHS Wordlist's 3rd band.*

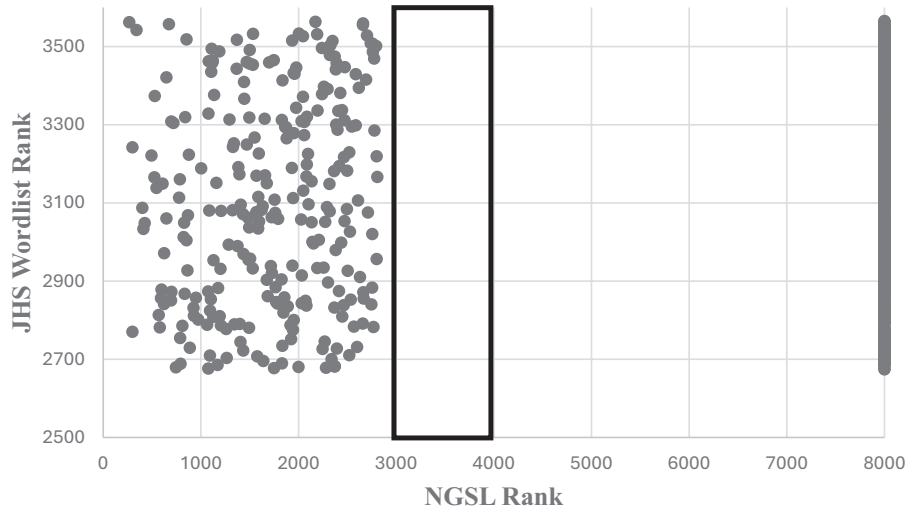


Figure 9 *Ranking comparisons between the NGSL and JHS Wordlist's 4th band.*

The final comparison between the NGSL and the compiled JHS wordlist illustrates an even less condensed pool of words, as shown in Figure 9. Considering the length of the NGSL, being a shorter list than the JHS wordlist, the items with difference values

in this comparison are exclusively off-listed items, at 70% of the total items within this band. 45 items (5%) are within the first 1000 words of the NGSL, with the remaining 25% items falling between the 1000 to 2800 range.

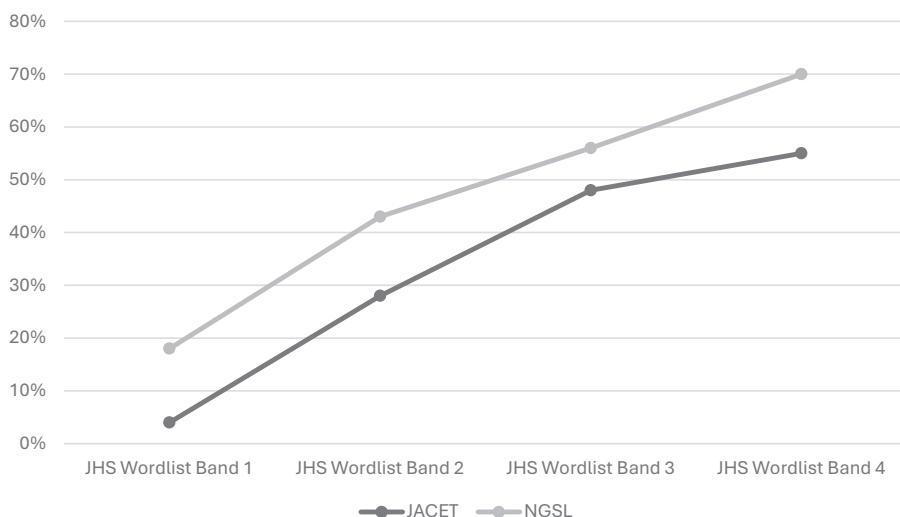


Figure 10 *Items per band with difference values more than 1000.*

The summation of these data points indicates that the first band is relatively aligned with both the JACET and NGSL lists in terms of the positions of individual words. In the case of both comparisons, the gap between rankings increases, mostly in part due to an increased number of items on the JHS wordlist that do not appear on either the NGSL or the JACET list. Naturally, as the NGSL is much shorter than the JACET list, the differences are larger as indicated in Figure 10. These differences only apply to words that are ranked higher on the JHS list than the JACET and NGSL, within their respective band. Thus, if one considers the first 3,000 lemmas to be high frequency words, there are a portion of words that are in separate bands but are still within the same frequency classification.

Discussion

The first research question aimed to examine the frequency and range of the items included on the JHS textbook wordlist, in order to take a broader look at the way JHS English textbooks introduce vocabulary to students, across the entire catalogue of MEXT approved textbooks. While there are plenty of shared themes, literature, and formats across textbooks, a large percentage of the items on the JHS wordlist are not universally covered across all textbooks, based on the range values reported in the analysis. The 41.8% of items with a reported range of two or less, is nearly identical to

the ranges reported by Hoshino (2020) for elementary textbooks, where 42% of items had a range of one. With only 802 items being shared across all the coursebook series, it is clear that vocabulary acquisition is not supported in the same way.

While the findings pertaining to range are unfavorable, there are some positives to be gleaned from the frequency values. As reported by Webb (2007), ten word repetitions contribute to gains in vocabulary acquisition, and with 40% of the items on the JHS wordlist having a frequency value greater than 10, approximately 1400 items are given adequate repetitions within the collective of JHS textbooks. Thus, while the range values deter the possibility of a core vocabulary shared by students across all textbooks, the frequency values support adequate encounter rates with a number of words relatively close to the MEXT goals for JHS students, granted that the words encountered would be different from coursebook to coursebook.

The purpose of the second and third research questions aimed to better understand more precisely what the content of the MEXT target words are, and how they might relate to large corpus-based lists such as the JACET8000 and NGSL. Particularly, these questions were posed to understand how prominently high frequency words featured within the catalogue of JHS textbooks in Japan, which are invaluable in language learning, especially reading, (Hunt & Beglar, 2005; Nation, 2023). The positives related to the frequency of items reported previously are also dependent upon how frequent those words are within the English language, and their value fluctuates depending on how useful they are as a result.

The main positive that can be taken from the results of these comparisons is that the first 1000 words of English, per both the JACET8000 and NGSL, is relatively well represented in the textbooks as a whole. Although the data has plenty of outliers, the difference of rank values indicates that the first band of the JHS wordlist is relatively aligned with the first band of the JACET8000 and NGSL. Given that the JHS wordlist was sorted by word power (frequency multiplied by range), approximately the first 900 items of the JHS wordlist are adequately encountered by JHS students, regardless of what textbook their schools are utilizing. This could indicate that incoming university students are more likely to have mastery of the first 1000 words compared to previous findings (e.g., McLean et al., 2014).

The analysis from the second band begins to illustrate a deviation from both the JACET and NGSL wordlists. As these results are not absolute values of the differences, the percentages shown are exclusive to items that are ranked lower on the established JACET and NGSL lists, and at a minimum are 1000 rankings lower than on the JHS wordlist. The implication of this shift in placement is an increase of mid-frequency or off listed words that appear in the JHS English textbooks. With that being said, there is also a portion of words that are still high-frequency words, being within the first 3000 words, albeit positioned differently from the JACET or NGSL rankings. An argument could be made that ultimately, the order of these words is not as important as their general inclusion within the textbooks, with the caveat being that those items ranked lower on the JHS wordlist, particularly from the third band onward, have low repetitions and ranges.

Within that vein, the point of discussion regarding the interpretation of such data, is not an endorsement of an absolute order to vocabulary acquisition, but rather, a wide view

of how textbooks should be utilized to support vocabulary acquisition, or other language related skills, such as intensive and extensive reading. There are several cases within the JHS wordlist that present words that have large difference values compared to the JACET ranking or NGSL ranking yet are not necessarily exclusively difficult or challenging. In the absence of a targeted vocabulary assessment for these words, it is inaccurate to suggest that the entirety of these items introduced to JHS students is in the wrong order, or, as previously mentioned, that the order based on encounter rate is the end all be all.

A textbook or supplemental learning material that is based on vocabulary levels from an established wordlist such as the JACET list or NGSL, could assume that a learner at a certain level should be able to read the targeted text with a certain degree of fluency. This analysis of JHS textbooks supports the idea that learners entering high school are reaching benchmarks established by MEXT in terms of number, yet not necessarily in relation to content. An example of this appears in the literature review, as Iwamoto and Kramer (2020) found that high school students would have adequate coverage of entrance examinations by studying the NGSL, whereas Nakayama (2022a) found that senior high school textbooks have relatively poor coverage of the NGSL. In this instance, a student would need to supplement their vocabulary knowledge outside of the textbooks they are learning from.

As indicated by Hammett (2020), the selection of textbooks from the MEXT catalogue is rarely, if at all, based on the content of the vocabulary, and the necessity to supplement moral education through English coursebooks undoubtedly plays a factor in the topics included within the textbooks. Given the prominence of proper nouns and cultural-specific words which were excluded from this JHS wordlist analysis, and that it should not be presumed that such words are automatically acquired or are not impactful on reading (Nicklin et al., 2023), there are even more factors to consider outside of the data presented here, in relation to the influence that textbooks have on vocabulary acquisition for JHS students in Japan.

Limitations

While the overall data and rankings for words in this study are accurate, there are still some limitations that should be considered. The first relates to the inclusion and exclusion of certain words, which were decided based on clear parameters, albeit without relying on precedence. In ambiguous cases, such as with non-English non-Japanese words, the items were excluded from the list if they were also not found on the JACET list. Additionally, the decision to give preference to the JACET list over the more widely used NGSL, was made solely based on the length of each respective list. The differences between the two might be considered insignificant, however, it is worth noting that the JACET list has some inconsistencies in regard to classifying the part of speech.

Lastly, the decision to attribute a value of 8001 to off-listed words for the NGSL could be misinterpreted given that the value of those words in terms of ranking is not known. Particularly in reporting ranked correlation values. Due to this factor the difference values for off-listed items were not emphasized in the results section for the NGSL, and the decision to set the value at 8001 was for visual and comparative purposes only.

Conclusion

Ultimately, the data derived from the JHS wordlist compiled in this study suggests that there is consistent representation of the first 1000 words of English based on two well-researched wordlists. While an argument could be made for MEXT approved JHS textbooks to more properly align with the NGSL and JACET8000 wordlists, it would be preferable to instead improve the ranges of words, as well as improve their frequency, to ensure that JHS students entering high school have a more streamlined exposure to such words, and perhaps align more closely to standardized entrance examinations.

While the ideal proposition would be to uncouple English and moral education in the curriculum standards, it would be more practical and easier for teachers to rely on the comparative data this study provides, in order to supplement textbooks with targeted vocabulary activities and evaluations. As textbooks are frequently updated and improved, it is important for them to be supported through evidence and applicable data. The purpose of this study is embedded in the idea that the target vocabulary suggested by textbooks are not necessarily vital for the greater language journey of Japanese students, and that, as suggested by others, teachers across Japan could vastly improve their students' vocabulary acquisition by focusing on high-frequency words from well-researched wordlists instead.

References

- Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Browne, C. (2024). NGSL Profiler [Computer Software]. Tokyo, Japan: Available from <https://www.ngslprofiler.com>
- Browne, C., Culligan, B., & Phillips, J. (2013). The new general service list. Retrieved from <http://www.newgeneralservicelist.org>
- Hammett, D. A. (2020). English coursebook adoption processes for junior high schools in Japan. In P. Clements, A. Krause, & R. Gentry, *Teacher efficacy, learner agency*. JALT, 55–66. <https://doi.org/10.37546/JALTPCP2019-08>
- Hoshino, Y. (2020). Vocabulary range and characteristics of words appearing in elementary school English textbooks in Japan. *ARELE: Annual Review of English Language Education in Japan*, 31, 49–63.
- Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a foreign language*, 17(1), 23–59.
- Ishikawa, S. I. (2019). A reconsideration of the construct of “A vocabulary for Japanese learners of English”: A critical comparison of the JACET wordlists and new general service lists. *Vocabulary Learning and Instruction*, 8(1), 1–7.
- Iwamoto, M., & Kramer, B. (2020). NGSL coverage of the SHS entrance exam and the National Center Test. *The Language Teacher*, 44(5), 3–7. <https://doi.org/10.37546/JALTTLT44.5-1>
- Klinger, R. (2024). Vocabulary frequency and dispersion in Japanese junior high school EFL textbooks. *Vocabulary Learning and Instruction*, 13(2), 1–18. <https://doi.org/10.29140/vli.v13n2.1171>

- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. <https://doi.org/10.1002/tesq.329>
- Muraoka, R. (2010). *Chugaku kentei kyokasho de gakushu sareru goi, gakushu sarenai goi: nobe gosu, kotonari gosu, goi renji no shiten kara [Learned vocabulary and not-learned vocabulary in junior high school authorized textbooks: From the perspectives of tokens, types, and vocabulary range]*. *STEP Bulletin*, 22, 182–203.
- McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction*, 3(2), 47–55. <https://doi.org/10.7820/vli.v03.2.mclean.et.al>
- MEXT, 2024. Junior High School Textbook Index. Retrieved from https://www.mext.go.jp/content/20230414-mxt_kyokasyo02-000029065_2.pdf
- MEXT. (2019). 中学校学習指導要領 (平成 29 年告示) 解説: 外国語編 [Junior high school course of study (2019) for foreign language studies]. Retrieved from https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afldfile/2019/03/18/1387018_010.pdf
- Mikajiri, N. (2023). Quantitative analysis of lexical bundles used in spoken English and Japanese elementary and junior high school English textbooks. *Journal of Corpus-based Lexicology Studies*, 5, 1–23.
- Mizumoto, A. (2021). *New Word Level Checker* [Web application]. <https://nwlc.pythonanywhere.com/>
- Mizumoto, A. (2015). Langtest (Version 1.0) [Web application]. Retrieved from <http://langtest.jp>
- Nakayama, S. (2022a). A close examination of vocabulary in Japanese EFL textbooks. In P. Ferguson, & R. Derrah (Eds.), *Reflections and new perspectives*. JALT. <https://doi.org/10.37546/JALTPCP2021-24>
- Nakayama, S. (2022b). Vocabulary in Japanese EFL textbooks: A bidirectional coverage analysis. In *The IAFOR International Conference on Education – Hawaii 2022 Official Conference Proceedings* (pp. 2189–1036).
- Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press.
- Nicklin, C., Patterson, A., & McLean, S. (2023). Quantifying proper nouns' influence on L2 English learners' reading fluency. *Studies in Second Language Acquisition*, 45(4), 906–929. <https://doi.org/10.1017/S027226312200050X>
- Ota, H et al. (2021a). Here we go! English course 1. Tokyo: Mitsumura Toshio.
- Ota, H et al. (2021b). Here we go! English course 2. Tokyo: Mitsumura Toshio.
- Ota, H et al. (2021c). Here we go! English course 3. Tokyo: Mitsumura Toshio.
- 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. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching1. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Stewart, J., Vitta, J. P., Nicklin, C., McLean, S., Pinchbeck, G. G., & Kramer, B. (2022). The relationship between word difficulty and frequency: A response to Hashimoto (2021). *Language Assessment Quarterly*, 19(1), 90–101. <https://doi.org/10.1080/15434303.2021.1992629>

- Takahashi, S., Hardy, T., Negishi, M., Hidai, S., Matsuzawa, S., Saito, E., ... Parmenter, L. (2012). *New crown English series* (Vols. 1–3). Sanseido.
- Teng, M., & Xu, J. (2022). Pushing vocabulary knowledge from receptive to productive mastery: Effects of task type and repetition frequency. *Language Teaching Research*, 29(2), 588–606. <https://doi.org/10.1177/13621688221077028>
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
- Vilkaitė-Lozdienė, L., & Schmitt, N. (2019). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In *The Routledge handbook of vocabulary studies* (pp. 81–96). Routledge.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130–163.
- Wongsarnpigoon, I. (2018). Vocabulary in junior high school textbooks and exams. In P. Clements, A. Krause, & P. Bennett (Eds.), *Language teaching in a global age: Shaping the classroom, shaping the world*. JALT.
- Zhang, P. (2022). How does repetition affect vocabulary learning through listening to the teacher's explicit instruction? The moderating role of listening proficiency and preexisting vocabulary knowledge. *Language Teaching Research*, 24(1). <https://doi.org/10.1177/13621688221140521>