Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2022, Volume 26, Issue 1
pp. 1–21

ARTICLE

Computing curriculum time and input for incidentally learning academic vocabulary

Clarence Green, *Federation University Australia*

Abstract

This paper computes estimates of the potential for Extensive Reading (ER) and Extensive Viewing (EV) to support the academic and discipline-specific vocabulary needs of students. While research into ER/EV for general vocabulary is well-established, only recently has academic vocabulary begun to be researched. Given curriculum time constraints, information on which academic vocabulary items might be learnable incidentally is useful, and this study provides teachers with information on which specific academic vocabulary items from multiple academic wordlists have a reasonable chance of being learned incidentally. It operationalizes ER/EV through corpora representing general fiction, television programs, and movies. It estimates the pedagogical time it would take to meet target vocabulary at different possible thresholds for incidental learning (6, 12, 20 times) with estimates for each computed for multiple possible reading rates (100, 260, 350 wpm) and viewing rates (80, 140, 200 wpm). Results report individual curriculum time/input estimates for over 2000 academic vocabulary targets across multiple subjects. Findings indicate ER/EV are pedagogies that could substantially support academic vocabulary development. A tool is released for teachers to compute personalized estimates using the reading rates of their students.

Keywords: *Extensive Reading, Extensive Viewing, Academic Vocabulary, Incidental Learning*

Language(s) Learned in This Study: *English*

APA Citation: Green, C. (2022). Computing curriculum time and input for incidentally learning academic vocabulary. Language Learning & Technology, 26(1), 1–21. https://doi.org/10125/73471

Introduction

This paper estimates computationally the potential for Extensive Reading (ER) and Extensive Viewing (EV) to support the academic and discipline-specific vocabulary needs of students. Research into ER/EV for general vocabulary within second language acquisition has been extensive (Webb, 2020), with the foundational paper from a computing perspective published in the current journal (Cobb, 2007). However, research into EV/ER to support academic and disciplinary vocabulary has only recently emerged. For example, Dang (2019) analyzed medical vocabulary in an EV corpus of medical television and McQuillan (2019) explored general academic vocabulary in an ER corpus of juvenile fiction. Both studies suggest significant amounts of academic/discipline-specific vocabulary are available from ER/EV, at rates potentially adequate for incidental learning within feasible pedagogical timeframes.

However, information that teachers/researchers do not yet have includes how much discipline-specific vocabulary is available in general ER/EV (i.e., corpora representing constructs of pleasure reading/viewing), an important component of most ER/EV approaches. Specialized vocabulary, such as those found in Dang (2019), may be broadly available. If teachers had this information for multiple disciplines, they would have a better sense of the potential of ER/EV for supporting disciplinary vocabulary. Further, estimates of input quantity and curriculum time would be valuable for specific vocabulary targets at multiple reading/viewing rates. This information needs to be provided in an easily accessible tool with the possibility of adjusting input rates for the needs of particular students. Finally, a larger corpus study than previously researched is required to not underestimate/overestimate the presence of academic

vocabulary through sampling bias, which is a possible limitation of McQuillan's (2019) analysis of academic vocabulary in a corpus of 22 novels (1 million words).

The following study therefore contributes a large-scale analysis of the Academic Word List (AWL) and Secondary Vocabulary Lists (SVL) in corpora representing possible ER/EV constructs, reporting which academic words are present in the input, which are not, and which might have the potential to be acquired incidentally. Curriculum time estimates for both general academic vocabulary and specialized vocabulary (chemistry, biology, physics, mathematics) are computed for variable reading/viewing rates. This study packages this information as a resource for teachers that allows for estimates to be adjusted according to their students and contexts. The aim of this study is to help teachers in planning which vocabulary they might be able to target through ER/EV and which they may decide to target through alternative pedagogies.

Extensive Reading and Viewing as Vocabulary Pedagogies

Extensive Reading (ER) is a collection of pedagogical approaches characterized by allowing students time to read widely material of interest, without necessarily requiring assessment. The pedagogy is also known as pleasure reading, free self-selected reading and sustained silent reading (Krashen, 2020). Renandya and Jacobs (2016, p. 98) characterize ER as consisting of a large amount of reading over extended time, with a focus on general comprehension rather than vocabulary learning. The focus on comprehension leads ER researchers to emphasize reading material be at an appropriate level, possibly no more than 2-5% of the vocabulary unknown (Hiebert, 2020). To ensure comprehensible input, several strategies can be used, either graded readers (Nation, 2014), teacher-librarian assistance, or free self-selected reading where students choose any text they want but are encouraged to put down anything too difficult or uninteresting and choose something else (Krashen, 2020). ER is a long-term curriculum commitment from weeks to years, and the longer students engage in ER, the more vocabulary is learned (Uchihara et al., 2019). The effectiveness of ER for vocabulary development is by now well established (Reynolds, 2020; Webb, 2020), and it is a rich pedagogy that supports literacy goals such as the enjoyment of literature and what this offers in terms of socio-cultural and cognitive development.

Extensive Viewing (EV) is a newer field of research/pedagogy that carries over much of the theory from ER. Webb (2020, p. 227) notes that as with ER, the current evidence is that EV effectively promotes vocabulary gains through incidental learning. Renandya and Jacobs (2016, p. 98) consider EV a type of extensive listening given the input is speech, but with additional benefits of visual support. EV, as operationalized by Webb (2015), "involves regular silent uninterrupted viewing of L2 television inside and outside of the classroom… including L2 movies… to improve vocabulary learning through increasing the amount of meaning-focused spoken input" (p. 161). EV can be enhanced by subtitles/closed captions (Peters, 2019). Understanding more about EV as a source of academic vocabulary input is important, given that Webb (2015, p. 159) cites figures for the average hours spent watching television ranging from 2.43 a day in Sweden to 8.18 hours in the United States. For reluctant readers, EV may be an appropriate scaffolding pedagogy if it can be shown to contain repeated encounters with academic vocabulary.

Incidental Vocabulary Learning and Repetition

Incidental vocabulary learning refers to vocabulary acquired within a meaning-focused task (e.g., reading comprehension) in which vocabulary learning is a by-product rather than the conscious focus of attention. Fundamental to incidental learning is repetition (Nation, 2014) and, while no number of encounters guarantees acquisition (Webb, 2020, p. 225), there has been extensive research exploring how many times target vocabulary may need to be encountered for incidental learning. A foundational study by Webb (2007) involved 121 L2 participants across four experimental groups. One experimental group encountered target words once, a second group 3 times, a third 7 times, and a fourth 10 times. Results indicated a general relationship between the number of repetitions and performance when tested, with significant incidental learning at 10 repetitions–a threshold at which students potentially can acquire "spelling, meaning, part of speech, words that it is associated with… [and] grammatical accuracy" (Webb, 2007, p. 62).

Vidal (2011) considered both reading and listening, taking three academic texts and lectures as input and testing the retention of 36 vocabulary targets at variable frequencies in a pretest-posttest-delayed posttest design. Participants included 248 first year undergraduates in an ESP tourism course across reading, listening and control groups. Vidal (2011, p. 247) reports regressions indicating, for reading, significant retention gains after three encounters with consistent retention increases from three to five repetitions, but no significant differences between five to six encounters. In the listening condition, retention gains started from three to five repetitions, with significant differences between five to six encounters. Overall, the study suggests participants needed higher rates of exposure for incidental vocabulary learning through listening than reading.

A recent meta-analysis by Uchihara et al. (2019) reviews multiple research estimates of the number of repetitions associated with incidental learning, including variations in these estimates depending on the student cohort and depth of vocabulary knowledge targeted. For native speakers, ten is cited as a possible threshold, while for L2 cohorts, estimates for form-meaning mappings range from two to four repetitions, eight to ten for more depth, and from 10 to 17 for associations and collocations. 12 encounters is noted as an oft-cited general estimate. Results of the meta-analysis indicate that incidental learning of vocabulary is probabilistic, with estimates ranging from a 36% to 54% chance of success, and no number of encounters consistently associated with acquisition. Nevertheless, the relationship between repetition and incidental learning is highly significant (r = .34) with mode of input (i.e., reading, viewing, listening) all significantly associated with acquisition at 0.41, 0.22 and 0.39 respectively. In sum, while quasi-experimental research has found certain repetition thresholds that result in more incidental learning than others, and these numbers are useful (Nation, 2014) for research and curriculum planning, no magic number can be given to teachers for all students, all contexts, nor all vocabulary.

Corpus Approaches to ER/EV as Input for Incidental Vocabulary Learning

Using estimates such as those cited above, a stream of research in corpus linguistics has involved computing the amount of input required from ER/EV for the number of encounters possibly needed for incidental vocabulary learning. The theoretical framework often being tested in corpus ER/EV research has been the pleasure reading (or viewing) hypothesis, situated within the broader framework of compelling comprehensible input (Krashen, 2020). A goal of researchers is to determine how far proficiency might develop by students self-selecting input that interests them. The research debate centers on the efficiency, efficacy and extent of incidental vocabulary learning compared to pedagogical options such as direct instruction (McQuillan, 2019).

The foundational study in this area is Cobb (2007), published in the current journal. He computed the frequencies of the first 3000 vocabulary items in the British National Corpus (BNC) in multiple possible representations of ER (175,000 words of general fiction; 163,000 words of newspapers; 179,000 words of academic texts) in order to estimate whether ER could provide this vocabulary frequently enough for incidental learning. Cobb (2007, p. 40) randomly selected 20 vocabulary samples from the 1000, 2000, and 3000 BNC word frequency bands, computing their frequency in the target corpora using six occurrences as the minimal threshold for incidental learning. Finding numerous vocabulary targets below six repetitions, Cobb (2007) concludes that explicit vocabulary teaching should supplement ER because the input required from ER to meet the incidental learning threshold would be too great, and hence the pedagogy would take too long to be a plausible solution to providing the vocabulary students need.

Nation (2014) states from his prior research that 98% of vocabulary in a text needs to be known for comprehension and that the first 9000 word families of English provides this coverage. He computes how much ER would be required to develop this vocabulary size incidentally. Nation sets an incidental learning threshold of 12 encounters as possibly "enough repetition … for learning to occur" (p. 12) and uses 25 novels from Project Gutenberg as a proxy for ER. In this corpus, he reports that from the 1000-2000 frequency band of the BNC, 805 targets would be encountered at least 12 times with approximately 171,411 words of input, equivalent to about two novels. Reading about three novels more with an additional 300,219 words of input would provide 830 targets from the 2-3000 frequency band, and so forth until after

approximately 2,956,908 words of input or 25 novels, students would have encountered 80% of the first 9000 words of the BNC at least 12 times, which Nation concludes may be "often enough to have a chance of learning them" (p. 7). He then asks if this amount of input is possible, computing how long it would take with reading rates of 150 and 200 words per minute. He reports that with about six hours per week, 40 weeks a year, or one to two hours of reading a day, it is possible to acquire the vocabulary needed for general comprehension within three to six years.

Researchers have used different reading, speech and listening rates in wpm for estimating curriculum time. As noted, Nation (2014) used a reading rate of 150 and 200 wpm. The current estimate for average reading rates amongst proficient adults based on a meta-analysis is 260 wpm with more advanced readers potentially reading up to 350 wpm (Brysbaert, 2019). In the study that follows, to provide useful information for slower readers, average readers and more proficient readers, three reading rates are computed providing curriculum time estimates for 100, 260 and 350 wpm. For EV, while the current best estimate of the average speech rate is 120-180 wpm (Brysbaert, 2019), there is significant variation reported in TV programs and movies. Nation (2014, p. 9) suggests a two-hour movie with 10,000 tokens equates to 83 words per minute. Spanos and Smith (1990) report the closed caption rate for Sesame Street is 60 wpm, Reading Rainbow 120 wpm, and ABC Evening News 250 wpm. Romero-Fresco (2016, p. 61) reports an average 141 wpm with a range from 74-231 wpm. In the present study, EV curriculum time estimates are therefore computed for three possible input rates approximating the range reported in previous research, namely 80, 140 and 200 wpm.

Corpus Approaches to ER/EV as Input for Academic Vocabulary

Research has begun to emerge considering the potential for ER/EV to facilitate the incidental learning of academic vocabulary. Academic vocabulary can be divided into general academic vocabulary that is widespread across subjects and discipline-specific vocabulary associated with particular subjects (Green, 2020). The extent to which these are mastered has an impact on student success, so it is an important research and pedagogical question as to how best to ensure the acquisition of this vocabulary (Crosthwaite & Cheung, 2019).

The Academic Word List (AWL) (Coxhead, 2000) has been the most widely used and researched representation of general academic vocabulary (Dang, 2020, p. 289). McQuillan (2019) compiled a 1,025,943-word ER corpus from 22 novels of recent, popular fiction (e.g., *Twilight*) and computed occurrences of AWL vocabulary in this data. He reports that 484 of 570 (85%) AWL words occur in his ER corpus, indicating that ER could support incidental learning of the AWL as it provides substantial exposure to general academic vocabulary. McQuillan (2019) then computes which AWL targets occur at multiple potential thresholds for incidental learning, namely 10, 12 and 25 recurrences in the ER corpus. McQuillan's (2019) estimates are that between 113 (20%) to 213 (37%) AWL targets could be acquired incidentally through ER, depending on the number of encounters targeted.

Focusing on discipline-specific vocabulary, Dang (2019) explored the potential for incidentally learning medical vocabulary through extensive viewing of medical television dramas. Dang (2019) developed a wordlist of 895 vocabulary targets from university lectures/seminars. Then, in a corpus of medical television (11,036,771 words), computed which would be encountered through EV and how frequently they occurred per episode, per season, all seasons and overall in the corpus. Using multiple possible thresholds for incidental learning, namely above 5, 10, 15, or 20, Dang reports "great potential for incidental learning" (p. 16), such as 76.26% of the specialized vocabulary encountered at least 5 times by watching multiple seasons of a program, 66.63% encountered 10 times, 60.73% 15 times, and 56.74% 20 times. If all seasons of all 37 programs in the corpus were viewed, 99.44% of the medical vocabulary would be attested 20 times or more.

For comparability with McQuillan (2019) and because it is so widely used (Dang, 2020) (making the results of this study of interest to a wide number of teachers), the AWL represents general academic vocabulary in the current study. Discipline-specific vocabulary is represented by the Secondary Vocabulary Lists (SVL), with 880 vocabulary targets for Biology, 519 for Chemistry, 253 for Mathematics, and 546 for Physics. The

SVL is the most extensive and recent vocabulary profile available for multiple disciplines in an L2 context (Green & Lambert, 2018). An added benefit of considering disciplinary vocabulary from the secondary/pre-tertiary level is that Uchihara et al. (2019, p. 584) find particularly strong incidental learning effects in secondary school, making it an educational context particularly worth exploring in relation to ER/EV.

The following study explores vocabulary profiled as useful for L2 secondary/pre-tertiary students, and is situated within L2 ER/EV research. However, an exciting aspect of ER research is that it has implications for many educational contexts (Green, 2020). Krashen (2004, 2020) has long argued that case histories, quasi-experimental and corpus research demonstrate pleasure-reading supports both monolingual and bilingual vocabulary, students in mainstream schools as well as in ESL or EFL contexts, and K-12 and adult learners. There is also methodological cross-pollination. Hiebert (2020) and Graves et al. (2019), prominent educational researchers in early literacy, have recently drawn on the methods of L2 corpus vocabulary research in ER to compute vocabulary recurrences and profiles of reading material in primary school, arguing for ER as a pedagogy to provide an essential early reading vocabulary.

This Study

Gallagher et al. (2019) note, "consideration of whether instruction is explicit or incidental is important given the time constraints teachers face when integrating content with literacy practices that are essential for all students" (p. 100). This study aims to provide teachers with information on which academic vocabulary items have a chance of being learned incidentally through ER/EV. It estimates the time it would take to meet the target vocabulary from the AWL and multiple disciplines in the SVL (chemistry, biology, physics, mathematics) at different possible thresholds for incidental learning, namely 6, 12 and 20 times. Estimates for each of these are computed for multiple possible reading rates, namely 100, 260, 350 wpm, and viewing rates, 80, 140, 200 wpm. The study does not aim at testing EV/ER effectiveness in comparison with explicit instruction, an important question but beyond the scope of this study. The contribution of this study is to examine what AWL/SVL words can be met by reading and viewing authentic materials, and how often.

This study adds important information to recent work on academic vocabulary in ER/EV. McQuillan's (2019) study of the AWL sampled a small ER corpus, possibly resulting in underestimates/overestimates of how much of the AWL is accessible through fiction. Teachers/researchers also do not have equivalent information on AWL affordances within EV. Dang (2019) notes "it is crucial to identify potential resources for incidental learning of specialized vocabulary" (p. 2) and this motivates her study of medical vocabulary in medical television. Yet discipline-specific vocabulary may be more broadly available. Potential resources could include general fiction, television and movies. Importantly, since a wide range of disciplines have not been studied, teachers do not have a broad picture of how far ER/EV can support the core vocabulary of multiple subjects.

Input should be compelling (Krashen, 2020) and this is why Dang (2020) examined medical dramas as a possible source of input for medical vocabulary, arguing medical professionals would presumably find it compelling. Therefore, if medical vocabulary frequently occurs in such television, then EV incorporating this material could be pedagogically useful. Yet as Krashen (2020) notes, current pedagogical advice to teachers is that informational and non-fiction texts be used to develop students' academic language, when it is much more likely that allowing them to choose fiction that interests them would be compelling. Research is therefore needed to evaluate if fiction contains the specialized vocabulary of academics, and at what rates.

This study focuses on the following two research questions:

1) Which academic vocabulary from the AWL and SVL are attested (i.e., present) in ER/EV input, and which are not?

2) For attested vocabulary, how much input would be required to encounter them 6, 12 and 20 times, and how long would this take for ER input at 100, 260, 350 wpm and EV input at 80, 140, 200 wpm?

The supplementary materials that accompany this study contain the further contribution of a tool for teachers, available here, that lists the individual ER/EV curriculum time/input estimates of over 2000 academic words from the AWL and SVL. Teachers can select, based on the profile of their students and their available curriculum time, which specific vocabulary items they might target through ER/EV. The tool, in the form of spreadsheets, also allows teachers to compute their own estimates using the reading rates of their students. There are simple online tools for computing reading rates in wpm, and the columns in the excel sheet are functions (e.g., [=H2/260/60]) where the middle value is the wpm reading rate. Teachers can change this value to the reading rate of their students and after copy and pasting the cell down the column, have personalized curriculum time estimates for every academic word in this study. The target number of encounters can also be altered by changing the initial value in the function (e.g., [=12/G2*1000000]) which computes curriculum time for 12 repetitions.

Methodology

To represent the vocabulary input possible from ER, this study built a larger corpus than previous studies (Cobb, 2007; McQuillan, 2019; Nation, 2014), representing the construct as a virtual library of books from which students might choose, and for EV a wide variety of television and movies. The corpora were designed to be an extensive sample, so as to be consistent with an ER/EV approach that emphasizes self-selection and putting down any book (or changing the channel) if the material is uninteresting or too difficult (Krashen, 2004). The self-selection component of extensive reading is important as it is a mechanism that allows students to navigate through authentic materials that are compelling and comprehensible. If students are encouraged to put down what they do not enjoy and do not comprehend, then as McQuillan (2016) shows, authentic fiction provides 'pathways' for students with smaller vocabularies to choose books at their level and progress over time to books where a larger vocabulary would be needed. Essential to this ER approach is that students have access to a wide range of books (Krashen, 2020). The large corpus of fiction books sampled in the current study was designed to represent this construct of a 'library' of options.

From the Corpus of Contemporary American English/COCA (Davies, 2010), the 5407 first chapters of novels representing its general fiction component were extracted using some ad-hoc python code. The titles of all novels are available in the supplementary materials here[1]. These titles suggest alignment with the principle of pleasure reading (e.g., *The Da Vinci Code*, *Diary of a South Beach Party Girl* rather than more 'high-brow' literature). The data constituted 21,226,768 words, henceforth referred to as the ER Corpus. An EV Corpus was also developed from COCA constituting 33,045,952 words from 369 movies, 2726 Fox and 5877 CNN television broadcasts. This EV corpus represented a wide-range of genres and program types, including *Point Break*, *Buffy the Vampire Slayer*, *Mission Impossible*, *Fox News*, *Hannity*, and *CNN Showbiz*. All titles included in the EV corpus are available in the supplementary materials here.

Corpora were lemmatized using the Natural Language Toolkit/NLTK (Bird et al., 2009). Frequency lists from the corpora were then generated using Wordsmith v7 (Scott, 2017) and matched against the AWL, biology, chemistry, physics and mathematics wordlists. For the SVL vocabulary targets, flemmas were matched, meaning words with the same form were counted together regardless of part of speech (Dang, 2020, p. 292). For the AWL, all word family members in the corpora were converted to the AWL headword form and then the ER/EV wordlists from Wordsmith matched against AWL headwords. To estimate the time needed to encounter target vocabulary at rates potentially sufficient for incidental learning, excel functions normed each target's frequency (per million words), then estimates were computed for the number of words of input required for the target number of encounters by dividing 6, 12, 20 by the normed figure and multiplying by 1 million. Using these estimates, reading/viewing times in hours were computed based on variable wpm input rates of 100, 260, 350 for reading and 80, 140, 200 for viewing, simply by dividing the input words by wpm and by 60 (minutes in an hour).

Results

This section first reports the results for general academic vocabulary represented by the AWL, and then the discipline-specific vocabulary from the SVL.

General Academic Vocabulary: The Academic Word List

Analysis of the AWL in the ER/EV corpora found all 570 targets attested. This differs from McQuillan (2019), whose ER corpus contained only 484/570, likely reflecting his smaller sample. Nevertheless, as with McQuillan (2019), what matters is not only attestation but whether the words are encountered at frequencies possibly sufficient for incidental learning. Table 1 illustrates the findings with curriculum time estimates for some AWL targets at multiple reading rates. The targets in the tables that follow are a few selections from amongst the most/least encountered academic vocabulary in ER/EV, as well as examples occurring in-between (e.g., approximately 12 times within a year). The complete resource (Appendix A) lists all vocabulary and their input/time estimates. The table is to be read as input size (in words), followed by time estimates (in hours) for the different wpm rates, as in Table 1. For example, to encounter "equivalent" 12 times would take 1.6 m (million) words of input, and 267/102/76 hours at reading rates of 100/260/350 wpm respectively.

Table 1 shows AWL vocabulary such as "found," "final," "job," "couple" and "respond" have a high possibility of incidental learning as they are easily available in ER. For as many as 20 encounters, a student reading at 260 wpm would only need approximately 10 hours of reading or less. Given this information, a teacher may decide to use ER for this vocabulary. On the other hand, Table 1 shows that words unlikely to be learned through ER, given the time and input required, include "empirical," "concurrent," "aggregate," and "qualitative." Yet, reflected in Table 1 is that for even the slowest reading rate, 20 encounters of words such as "ethic," "equivalent," "statistic," "subsequent," and "amend" could be achieved with 45 minutes of reading over 6 to 18 months. A significant overall result (Appendix A) is that with 45 minutes of ER over three years, 562 of 570 AWL words would be met at least 12 times by students reading about 260 wpm, equivalent to 98.5% of the AWL. Even if the probability of acquisition is a proportion of this, such findings are promising for the potential for ER to carry some of the pedagogical burden of providing academic vocabulary.

The AWL results from the EV corpus are reported in Table 2. Table 2 indicates EV provides a significant number of AWL encounters within reasonable timeframes. For example, "job," "issue," and "economy" are encountered multiple times within a matter of hours, and vocabulary such as "protocol," "external," and "confine" would all be encountered at least 12 times within a year of less than an hour of EV per day, at all but the slowest input rate. Supposing students watched one hour per day, overall results suggest that within one year students would meet 557/570 AWL words at least six times at the average input rate of 140 wpm–an impressive 97.7%. Estimating for the slowest input rate (80 wpm) after one year, approximately 432 (75.8%) of the AWL would have been encountered 20 times, including "contrary," "flexible," "scheme," "discrete," "diminish," and "parallel." This leaves only 138 AWL words a teacher may need to target through alternative pedagogies, such as those in final row of Table 2 (e.g., "concurrent," "empirical," "paradigm," and "interval").

Table 1

Estimates of ER Input, Reading Time for Repeated Encounters of AWL Words

AWL	6 Encounters Input Size (words): 100, 260, 350wpm (hours)	12 Encounters Input Size (words): 100, 260, 350wpm (hours)	20 Encounters Input Size (words): 100, 260, 350 wpm (hours)
found	11516: 1.92, 0.74, 0.53	23033: 3.84, 1.48, 1.10	38388: 6.40, 2.46, 1.83
final	16165: 2.69, 1.04, 0.75	32329: 5.39, 2.07, 1.54	53882: 8.98, 3.45, 2.57
job	18815: 3.14, 1.21, 0.87	37631: 6.27, 2.41, 1.79	62718: 10.45, 4.02, 2.99
couple	26533: 4.42, 1.70, 1.23	53067: 8.84, 3.40, 2.53	88445: 14.74, 5.67, 4.21
respond	42839: 7.14, 2.75, 1.98	85678: 14.28, 5.49, 4.08	142797: 23.80, 9.15, 6.80
shift	48518: 8.09, 3.11, 2.25	97037: 16.17, 6.22, 4.62	161728; 26.95, 10.37, 7.70
migrate	786177: 131, 50, 36	1.57m: 262, 100, 74	2.62m: 436, 167, 124
ethic	791060: 131, 50, 36	1.58m: 263, 101, 75	2.63m: 439, 169, 125
equivalent	801010: 133, 51, 37	1.60m: 267, 102, 76	2.67m: 445, 171, 127
statistic	816414: 136, 52, 37	1.63m: 272, 104, 77	2.72m: 453, 174, 129
subsequent	821681: 136, 52, 38	1.64m: 273, 105, 78	2.73m: 456, 175, 130
amend	827017: 137, 53, 38	1.65m: 275, 106, 78	2.75m: 459, 176, 131
parameter	6.3m: 1061, 408, 294	12.7m: 2122, 816, 606	21.2m: 3537, 1360, 1010
intermediate	7.4m: 1248, 480, 346	14.9m: 2497, 960, 713	24.9m: 4162, 1600, 1189
allocate	8.4m: 1415, 544, 393	16.9m: 2830, 1088, 808	28.3m: 4717, 1814, 1347
concurrent	31.8m: 5306, 2041, 1474	63.6m: 10613, 4082, 3032	106.1m: 17688, 6803, 5053
aggregate	42.4m: 7075, 2721, 1965	84.9m: 14151, 5442, 4043	141.5m: 23585, 9071, 6738
qualitative	127m: 21226, 8164, 5896	25.4m: 42453, 16328, 12129	424.5m: 70755, 27213, 20215

Table 2

Estimates of EV Input, Viewing Time for Repeated Encounters of AWL Words

AWL	6 Encounters Input Size (words): 80, 140, 200wpm (hours)	12 Encounters Input Size (words): 80, 140, 200wpm (hours)	20 Encounters Input Size (words): 80, 140, 200wpm (hours)
issue	9338: 1.95, 1.11, 0.78	18675: 3.89, 2.22, 1.56	31126: 6.48, 3.71, 2.59
percent	12603: 2.63, 1.50, 1.05	25207: 5.25, 3.00, 2.10	42011: 8.75, 5.00, 3.50
job	12822: 2.67, 1.53, 1.07	25644: 5.34, 3.05, 2.14	42739: 8.90, 5.09, 3.56
secure	13420: 2.80, 1.60, 1.12	26839: 5.59, 3.20, 2.24	44732: 9.32, 5.33, 3.73
economy	16450: 3.43, 1.96, 1.37	32901: 6.85, 3.92, 2.74	54834: 11.42, 6.53, 4.57
area	19070: 3.97, 2.27, 1.59	38141: 7.95, 4.54, 3.18	63568: 13.24, 7.57, 5.30
forthcoming	1011611: 210, 120, 84	2023222: 421, 240, 168	3372036: 702, 401, 281
input	1016799: 211, 121, 84	2033597: 423, 242, 169	3389328: 706, 403, 282
protocol	1022040: 212, 121, 85	2044080: 425, 243, 170	3406799: 709, 405, 283
confine	1038093: 216, 123, 86	2076185: 432, 247, 173	3460309: 720, 411, 288
external	1120202: 233, 133, 93	2240404: 466, 266, 186	3734006: 777, 444, 311
commodity	1180213: 245, 140, 98	2360425: 491, 281, 196	3934042: 819, 468, 327
paradigm	4.0m: 843, 481, 337	8.1m: 1686, 963, 674	13.4m: 2810, 1605, 1124
intermediate	4.3m: 897, 513, 359	8.6m: 1795, 1026, 718	14.3m: 2993, 1710, 1197
deviate	4.4m: 917, 524, 367	8.8m: 1835, 1049, 734	14.7m: 3059, 1748, 1223
interval	5.2m: 1087, 621, 434	10.4m: 2174, 1242, 869	17.4m: 3623, 2070, 1449
empirical	5.6m: 1180, 674, 472	11.3m: 2360, 1348, 944	18.9m: 3934, 2248, 1573
concurrent	6.6m: 1376, 786, 550	13.2m: 2753, 1573, 1101	22.0m: 4589, 2622, 1835

Discipline-Specific Academic Vocabulary: Biology

The SVL contains 857 biology words, of which 680 were attested in the ER corpus and 683 in the EV corpus. The substantial majority of the SVL: biology (79-80%) is therefore present in ER/EV, though in comparison with the AWL, biology vocabulary seems less available, with 177 targets unattested. Unattested vocabulary include "homozygous," "genotype," "cytoplasm," and "chromatid" (the complete list for all subjects is provided here). For attested vocabulary, Table 3 reports selected examples of curriculum time estimates.

Table 3

Estimates of ER Input, Reading Time for Repeated Encounters of Biology Words

SVL: Biology	**6 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**12 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**20 Encounters Input Size (words): 100, 260, 350wpm (hours)**
eye	3850: 0.64, 0.25, 0.18	7699: 1.28, 0.49, 0.37	12832: 2.14, 0.82, 0.61
hair	9247: 1.54, 0.59, 0.43	18494: 3.08, 1.19, 0.88	30824: 5.14, 1.98, 1.47
light	9462: 1.58, 0.61, 0.44	18924: 3.15, 1.21, 0.90	31541: 5.26, 2.02, 1.50
pull	10872: 1.81, 0.70, 0.50	21743: 3.62, 1.39, 1.04	36239: 6.04, 2.32, 1.73
body	12814: 2.14, 0.82, 0.59	25628: 4.27, 1.64, 1.22	42714: 7.12, 2.74, 2.03
suspension	1872950: 312, 120, 86	3745900: 624, 240, 178	6243167: 1040, 400, 297
syndrome	1990010: 331, 127, 92	3980019: 663, 255, 189	6633365: 1105, 425, 315
stamen	2122677: 353, 136, 98	4245354: 707, 272, 202	7075589: 1179, 453, 336
stimulate	2158654: 359, 138, 99	4317309: 719, 276, 205	7195515: 1199, 461, 342
modify	2195873: 365, 140, 101	4391745: 731, 281, 209	7319575: 1219, 469, 348
apparatus	2234397: 372, 143, 103	4468793: 744, 286, 212	7447989: 1241, 477, 354
blindness	2234397: 372, 143, 103	4468793: 744, 286, 212	7447989: 1241, 477, 354
germination	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
homologous	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
hydrostatic	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
lactic	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
lumen	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
maltose	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
mutated	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215

Table 3 indicates biology-relevant vocabulary such as “eye,” “body,” “hair” are easily available from ER, while vocabulary less available includes “homologous,” “hydrostatic,” “lumen,” with estimates of approximately 73 years of reading at 45 minutes a day needed to meet them 6 times for a slow reader. However, vocabulary encountered at least 12 times a year with 45 minutes a day of reading at input rates above 100 wpm include “syndrome,” “stamen,” “modify,” “stimulate.” Even for a slow reader at 100 wpm, overall results are that 45 minutes of ER a day could provide 12 encounters with up to 243 biology targets within a year, 319 after two, and 365 after three. Reading at 260 wpm, estimates are 336 words after one year, 419 after two, 464 after three; and for faster readers at 350 wpm, 387 words after one year, 454 after two, and 493 after three. Thus, for all reading rates, 47 to 57% of the target biology vocabulary might with some confidence be targeted within one to three years through ER. Table 4 illustrates the findings for EV.

Table 4

Estimates of EV Input, Viewing Time for Repeated Encounters of Biology Words

SVL: Biology	6 Encounters Input Size (words): 80, 140, 200wpm (hours)	12 Encounters Input Size (words): 80, 140, 200wpm (hours)	20 Encounters Input Size (words): 80, 140, 200wpm (hours)
eye	12838: 2.67, 1.53, 1.07	25677: 5.35, 3.06, 2.14	42795: 8.92, 5.09, 3.57
pull	14131: 2.94, 1.68, 1.18	28263: 5.89, 3.36, 2.36	47104: 9.81, 5.61, 3.93
light	19477: 4.06, 2.32, 1.62	38954: 8.12, 4.64, 3.25	64923; 13.53, 7.73, 5.41
inside	20517: 4.27, 2.44, 1.71	41034: 8.55, 4.88, 3.42	68390; 14.25, 8.14, 5.70
wall	21715: 4.52, 2.59, 1.81	43429; 9.05, 5.17, 3.62	72382; 15.08, 8.62, 6.03
artery	1279198: 266, 152, 106	2558396: 533, 304, 213	4263994: 888, 507, 355
modify	1304445: 271, 155, 108	2608891: 543, 310, 217	4348152: 905, 517, 362
embryonic	1313084: 273, 156, 109	2626168: 547, 312, 218	4376947: 911, 521, 364
suspension	1313084: 273, 156, 109	2626168: 547, 312, 218	4376947; 911, 521, 364
circular	1321838: 275, 157, 110	2643676: 550, 314, 220	4406127: 917, 524, 367
muscular	1348814: 281, 160, 112	2697629: 562, 321, 224	4496048: 936, 535, 374
coordination	1376915: 286, 163, 114	2753829: 573, 327, 229	4589716: 956, 546, 382
homeostasis	198m: 41307,23604,16522	396m: 82614,47208,33045	661m: 137691,78680, 55076
maize	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691,78680, 55076
mitotic	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691, 78680, 55076
ovum	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691, 78680, 55076
pharynx	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691, 78680, 55076
phenotype	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691, 78680, 55076
sclera	198m: 41307,23604,16522	396m: 82614, 47208, 33045	661m: 137691, 78680, 55076

Table 4 shows that the words "eye," "inside," and "wall" are likely to be acquired through EV (if not already known) with at least 20 encounters at all input rates achieved within 15 hours. Vocabulary such as "artery," "modify," "embryonic," "muscular" also appear good candidates for EV, occurring from 6–20 times at all input rates within one to three years of one hour a day of viewing. Given the time and input estimates, words unlikely to be acquired include "phenotype," "mitotic," "homeostasis" and so forth. Overall results indicate that with input at 140 wpm, targeting 12 encounters, after one year of one hour a day, up to 299 biology-relevant vocabulary may have been encountered enough to be incidentally learned. These include "evolution," "fetal," "protein," "nutrition"; after two years, up to 373 words, including "testosterone," "pesticide," "antibiotic," "marrow" and after three years, up to 423 words including "organism," "iris," and "toxin."

Discipline-Specific Academic Vocabulary: Chemistry, Physics and Mathematics

Given space limitations, results for chemistry, physics and mathematics are summarized together. Readers are referred to the complete results and the Appendix for some illustrative tables and examples discussed in this section. For chemistry vocabulary, 402 out of 501 were present in the ER corpus, and 396 present in the EV corpus. The majority of the chemistry wordlist (79–80%) was therefore in general ER/EV.

Unattested targets outside the scope of ER/EV include technical vocabulary such as "ethane," "stoichiometry." Appendix A shows other chemistry words unlikely to be acquired through ER, such as "oxidation," "silica," "valence," with estimates of 127–254 million words of input needed for 12 encounters at 260 wpm. However, results show that planning for 45 minutes a day, 197 chemistry targets would be encountered more than 12 times after one year at 260 wpm, including "liberate," "purity," "carbon," "vigorous," growing to 249 targets after two years, and 266 after three. For EV, Appendix B shows that words such as "displace," "collide," "silicon," "spontaneous" might be acquirable through EV, as all would be encountered at least 20 times within three years at all input rates with one hour of daily viewing. A curriculum planned on estimates of 140 wpm and 12 encounters could with some assurance place the first 169 chemistry words listed in the supplementary tool as having a chance of incidental learning within a year, 206 by the second and 231 by the third.

For physics, both ER/EV contained 496 out of 523 targets, which is 95% of the specialized wordlist. The handful of unattested vocabulary include "solenoid," "voltmeter," and "diffraction." Appendix C shows other low frequency items that would need supplementary pedagogy, such as "rarefaction," "oscillation." However, vocabulary such as "compress," "sensor," "fundamental," "sensitivity," "aerial," all seem possible targets for ER. At 45 minutes reading a day, they would be encountered within one year at least 12 times at 260 wpm. Within three years, up to 398 physics words (76% of the wordlist) occur more than 12 times, including "conical," "sparingly," "titanium." Results for EV indicate that at one hour per day, vocabulary such as "disc," "rotate," "precision," and "dense," are reasonable candidates for incidental learning. As Appendix D shows, they would be met (with the exception of the slowest input rate) 12 times within a year and 20 times within two years. Overall, with one hour of EV a day at input of 140 wpm, 317 targets would be encountered at least 12 times within one year, 372 within two years, and 381 within three years.

For mathematics, 232 out of 242 (96%) targets were in the ER corpus and 223 out of 242 (92%) in the EV corpus. Mathematics vocabulary with a high chance of incidental learning through ER, illustrated in Appendix E, include "width," "intercept," "radius," and "acute," which occur over 20 times within one to two years above input rates greater than 100 wpm. With 45 minutes of ER per day at 260 wpm, up to 133 (55%) of the mathematics wordlist would be encountered at least 20 times within a year, vocabulary including "function," "method," "sum," "solution," "arc," "index," and "significant." After three years, estimates reach up to 171 words (70%) more than 20 times, including "integer," "cosine," "quadrilateral," "gradient," and "exponential." For EV, vocabulary such as "probability," "domain," "integration," "axis," "cone," and "consecutive" might reasonably be placed under EV. Illustrated in Appendix F, they all occur at least 12 times within a year of one hour viewing a day. Up to 135 targets would be encountered at least 12 times after one year of EV assuming one hour per day at 140 wpm, 151 targets after two years, and 165 targets after three years.

General Discussion

This study estimated the potential for ER and EV to support the incidental learning of the general academic vocabulary of the AWL, and the discipline-specific academic vocabulary of the SVL. One hopes this study is useful in providing curriculum time estimates for specific academic vocabulary at different input rates and repetition thresholds, thereby helping teachers make curriculum planning decisions about which words they can with some confidence place under ER or EV pedagogy. Furthermore, the supplementary material of this study constitutes a tool in which teachers can enter their students' reading rates and compute personalized curriculum time estimates.

One clear and important finding is that ER and EV provide substantial exposure to both general academic and discipline-specific vocabulary. This suggests they are valuable pedagogies for supporting academic vocabulary development. The study could have potentially found that ER/EV would result in limited encounters with academic vocabulary, or that curriculum times to encounter this vocabulary frequently enough for possible incidental learning would be unreasonable, but this was not the case. Rather, for the

AWL, it was found that all 570 word families were present in both ER and EV corpora, and within three years at an average reading rate of 260 wpm, up to 98.5% of the AWL would have been encountered 12 times or more. Even at a slower reading rate of 100 wpm, a student reading for 45 minutes a day on average over three years would have encountered 529 of 570 (93%) AWL targets at least 12 times.

This study finds that the AWL is more prevalent in general fiction than the previous estimates of McQuillan (2019), likely due its larger corpus. Building on Dang (2019), this study shows that exposure to disciplinary vocabulary need not necessarily come from materials related to target disciplines but can be supported with pleasure reading/viewing generally. Results consistently indicated that approximately 40-60% of target vocabulary from mathematics, chemistry, physics and biology, may have sufficient encounters for incidental learning within one to three years of curriculum time. Since neither ER/EV are the only source of vocabulary input, additional items with fewer encounters may be also be learnable.

Being the first large-scale corpus study to simultaneously consider ER and EV, this study can also report the new finding that the vocabulary input from EV is not particularly impoverished compared to ER. This does not downplay the importance of having students read. Rather, it simply suggests that a direct comparison of ER/EV offers comparable overall rates of attested/unattested target vocabulary. For example, both ER and EV offered exposure to about 80% of the biology and chemistry wordlists, about 95% for physics, though for mathematics ER offered slightly more coverage (96%) than EV (92%). Another takeaway from these numbers is that at ER/EV are especially relevant input for physics and math.

Pedagogical Implications

Let us explore some suggestions for L2 teachers about how they can use the tool developed by this study. This tool is a spreadsheet with embedded functions. It lists input and curriculum time estimates for every target word on the AWL/SVL. To provide but one example of its pedagogical value, in the context of the principal researcher's experience as a high-school English/ESL teacher and curriculum coordinator for grades 7 to 10, the tool would be useful for long-term curriculum planning. ESL students coming into grade 7 have a foundational vocabulary but, like monolingual students, transition from a broader middle school curriculum to more specialized subjects over the next few years, with their final school years characterized by electives aligned with their future plans. If proficiency between grades 7 to 10 in academic/disciplinary vocabulary is not achieved, this contributes to students struggling in certain subjects, and influences their life-choices because students avoid choosing electives where they lack the confidence to succeed. Because this tool tells us which vocabulary has a good chance of coming from ER input within certain timeframes, we can help students by planning a curriculum that includes daily pleasure reading from grade 7 onward. Teachers might plan for 4 terms x 10 weeks, 45 minutes a day, which would be 150 hours of ER per year. Using this timeframe, one looks at the tool and finds that under 150 hours at the average reading rate, 341 AWL words have at least 12 exposures. It seems reasonable that instruction should support AWL vocabulary less likely to be incidentally learned. Vocabulary outside general ER/EV does not mean direct instruction is the only alternative. If a faculty can agree on an ER program all the way from grades 7 to 10, the years after this may focus on narrow reading matched to the interest of the students (Krashen, 2020).

Pre-tertiary foundation L2 teachers more generally can benefit from this tool as it allows them to make estimates for their available curriculum time by having students estimate their reading rate in wpm (online tools for this being easily available). Following the instructions in the methodology, teachers can enter this reading rate into the tool and compute an individual vocabulary profile for each student. This might, for example, inform how teachers balance ER and EV for particular students. Consider a student with literacy difficulties and who was a slow reader under 100 wpm (perhaps lacking motivation) yet watched input above 140 wpm with comprehension and enjoyment. The tool suggests that ER for this student at 45 minutes a day would provide less academic vocabulary exposure than one hour of EV per day. Teachers and literacy aides may decide based on this information to increase EV relative to ER in a differentiated curriculum for this student's needs.

When looking at the results of this study, teachers should bear in mind that how words are used in EV/ER

may differ from their use in academic contexts and specialized uses may need alternative instruction. For example, from mathematics, "mean" was frequently encountered in ER/EV, as was "work" from physics. However, this is somewhat misleading as in the context of mathematics "mean" is a synonym for "average," likely a minor meaning encountered in ER/EV. Similarly, "work" in physics refers to force acting on an object. This does not mean that ER/EV cannot sometimes provide specialized meanings, however, such as in this example in *The Hammer of God* by Arthur C Clarke (1993), "it's moving in a retrograde orbit- that is, in the opposite direction to all the planets" (p. 148), from which one could potentially establish a meaning for "retrograde" and "orbit."

Limitations and Future Research

No number of encounters guarantees incidental learning (Webb, 2020), only a probability of learning, and just because vocabulary is available in fiction does not mean that vocabulary support will not be needed during classroom interactions between teachers and students engaged with specific texts. Another issue with the results is the spacing issue (Uchihara et al., 2019). Obviously, there is a difference between any target numbers of encounters spaced over a month versus over a year. At this stage of ER/EV research, we do not know the right equation of time to repetitions so we cannot pass this on to teachers. We do not know, for example, whether 12 encounters over a year is too little or sufficient for incidental learning. Future research will have to fill this research gap.

Conclusion

Curriculum decisions about the academic vocabulary that students need and the most effective pedagogy for promoting the acquisition of this vocabulary depend on the profile of any given cohort of students. Teachers need to take account of many moderating variables when putting pedagogical research into practice, such as the language background of their students, their motivation, age, proficiency, whether they are struggling readers, training for professions, or aiming for general functional literacy. Having said that, there is a place for research that considers vocabulary curriculum planning more broadly. This current study found we may be able to partition out to ER/EV pedagogies some of the learning burden for general and discipline-specific academic vocabulary. Since this paper provides evidence that there is enough exposure in ER/EV for the potential acquisition of particular academic vocabulary items, we can incorporate this information into our curriculum planning for which academic words we might want to target with ER/EV and which through alternative pedagogies.

Notes

1. The novels listed are described in Green (2020) as being full-length within COCA, but in fact COCA samples their first chapters.

References

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media. https://www.nltk.org/book/

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Clarke, Arthur C. (1993). *The hammer of God*. Orbit.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38–63. http://dx.doi.org/10125/44117

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Crosthwaite, P., & Cheung, L. (2019). *Learning the language of dentistry: Disciplinary corpora in the teaching of English for Specific Academic Purposes* (Vol. 93). John Benjamins. https://doi.org/10.1075/scl.93

Dang, T. N. Y. (2019). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related tv programs: Insights from medical corpora. *TESOL Quarterly 54*(2), 436–459**.** https://doi.org/10.1002/tesq.552

Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 288–304). Routledge. https://doi.org/10.4324/9780429291586

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Gallagher, M. A., Barber, A. T., Beck, J. S., & Buehl, M. M. (2019). Academic vocabulary: Explicit and incidental instruction for students of diverse language backgrounds. *Reading & Writing Quarterly*, *35*(2), 84–102.

Graves, M. F., Elmore, J., & Fitzgerald, J. (2019). The vocabulary of core reading programs. *The Elementary School Journal*, *119*(3), 386–416.

Green, C. (2020). Extensive reading and viewing as input for academic vocabulary: A large-scale vocabulary profile coverage study of students' reading and writing across multiple secondary school subjects. *Lingua*, *239*, 102838. https://doi.org/10.1016/j.lingua.2020.102838

Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes*, *35*, 105–115. https://doi.org/10.1080/10573569.2018.1510796

Hiebert, E. H. (2020). The core vocabulary: The foundation of proficient comprehension. *The Reading Teacher*, *73*(6), 757–768. https://doi.org/10.1002/trtr.1894

Krashen, S. (2004). The case for narrow reading. *Language Magazine*, *3*(5), 17–19.

Krashen, S. (2020). Self-selected fiction: The path to academic success? *Synergy*, *18*(1), 1–2.

McQuillan, J. (2016). What can readers read after graded readers? *Reading in a Foreign Language*, *28*(1), 63–78. https://doi.org/10125/66715

McQuillan, J. (2019). Where do we get our academic vocabulary? Comparing the efficiency of direct instruction and free voluntary reading. *The Reading Matrix: An International Online Journal*, *19*(1), 129–138.

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16. https://doi.org/10125/66881

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Renandya, W. A., & Jacobs, G. M. (2016). Extensive reading and listening in the L2 classroom. In W. A. Renandya, & Handoyo, P. (Eds.), *English language teaching today* (pp. 97–110). Springer. https://doi.org/10.1007/978-3-319-38834-2

Reynolds, B. L. (2020). Situated incidental vocabulary acquisition: The effects of in-class and out-of-class novel reading. *Applied Linguistics Review*, (ahead-of-print). https://doi.org/10.1515/applirev-2019-0059

Romero-Fresco, P. (2016). Accessing communication: The quality of live subtitles in the UK. *Language & Communication*, *49*, 56–69. https://doi.org/10.1016/j.langcom.2016.06.001

Scott, M. (2016). *WordSmith tools* (version 7) [computer software]. Liverpool.

Spanos, G., & Smith, J. (1990). *Closed captioned television for adult LEP literacy learners* (ED321623). ERIC. https://eric.ed.gov/?id=ED321623

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

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language Learning Beyond the Classroom* (pp. 159–168). Routledge.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Appendix A

Estimates of ER Input, Reading Time for Repeated Encounters of Chemistry Words

SVL: Chemistry	**6 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**12 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**20 Encounters Input Size (words): 100, 260, 350wpm (hours)**
carbon	1369469: 228, 87, 63	2738938: 456, 175, 130	4564896: 760, 292, 217
vigorous	1369469: 228, 87, 63	2738938: 456, 175, 130	4564896: 760, 292, 217
poisonous	1592008: 265, 102, 73	3184015: 530, 204, 151	5306692: 884, 340, 252
liberate	1744666: 290, 111, 80	3489332: 581, 223, 166	5815553: 969, 372, 276
purity	1872950: 312, 120, 86	3745900: 624, 240, 178	6243167: 1040, 400, 297
polar	2054203: 342, 131, 95	4108407: 684, 263, 195	6847345: 1141, 438, 326
litmus	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
neutron	63.6m: 10613, 4082, 2948	12.7m: 21226, 8164, 6064	212m: 35377, 13606, 10107
nitrous	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
oxidation	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
silica	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
alkaline	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215

Appendix B

Estimates of EV Input, Viewing Time for Repeated Encounters of Chemistry Words

SVL: Chemistry	**6 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**12 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**20 Encounters Input Size (words): 100, 260, 350wpm (hours)**
displace	1313084: 273, 156, 109	2626168: 547, 312, 218	4.3m: 911, 521, 364
collide	1358053: 282, 161, 113	2716106: 565, 323, 226	4.5m: 943, 538, 377
aspirin	1426444: 297, 169, 118	2852888: 594, 339, 237	4.7m: 990, 566, 396
vigorous	1426444: 297, 169, 118	2852888: 594, 339, 237	4.7m: 990, 566, 396
apparatus	1447268: 301, 172, 120	2894536: 603, 344, 241	4.8m: 1005, 574, 402
silicon	1468709: 305, 174, 122	2937418: 611, 349, 244	4.8m: 1019, 582, 407
spontaneous	1490795: 310, 177, 124	2981590; 621, 354, 248	4.9m: 1035, 591, 414
fluorine	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
iodide	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
phosphoric	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
silicate	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
solute	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
strontium	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076

Appendix C

Estimates of ER Input, Reading Time for Repeated Encounters of Physics Words

SVL: Physics	**6 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**12 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**20 Encounters Input Size (words): 100, 260, 350wpm (hours)**
telescope	1415118: 235, 90, 65	2830236: 471, 181, 134	4717060: 786, 302, 224
compress	1447280: 241, 92, 67	2894559: 482, 185, 137	4824265: 804, 309, 229
sensor	1463915: 243, 93, 67	2927830: 487, 187, 139	4879717: 813, 312, 232
fundamental	1572353: 262, 100, 72	3144706: 524, 201, 149	5241177: 873, 335, 249
aerial	1698141: 283, 108, 78	3396283: 566, 217, 161	5660471: 943, 362, 269
radiation	1819437: 303, 116, 84	3638875: 606, 233, 173	6064791: 1010, 388, 288
generator	1872950: 312, 120, 86	3745900: 624, 240, 178	6243167: 1040, 400, 297
oscillation	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
polarity	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
radioactivity	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
transverse	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
deceleration	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
polystyrene	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
rarefaction	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215

Appendix D

Estimates of EV Input, Viewing Time for Repeated Encounters of Physics Words

SVL: Physics	**6 Encounters Input Size (words): 80, 140, 200wpm (hours)**	**12 Encounters Input Size (words): 80, 140, 200wpm (hours)**	**20 Encounters Input Size (words): 80, 140, 200wpm (hours)**
disc	1054658: 219, 125, 87	2109316: 439, 251, 175	3515527: 732, 418, 292
rotate	1159507: 241, 138, 96	2319014: 483, 276, 193	3865024: 805, 460, 322
frequency	1216415: 253, 144, 101	2432831: 506, 289, 202	4054718: 844, 482, 337
transparent	1223924: 254, 145, 101	2447848: 509, 291, 203	4079747: 849, 485, 339
precision	1304445: 271, 155, 108	2608891: 543, 310, 217	4348152: 905, 517, 362
dense	1348814: 281, 160, 112	2697629: 562, 321, 224	4496048: 936, 535, 374
collide	1358053: 282, 161, 113	2716106: 565, 323, 226	4526843: 943, 538, 377
propagation	99m: 20653, 11802, 8261	198m: 41307, 23604, 16522	330m: 68845, 39340, 27538
rarefaction	99m: 20653, 11802, 8261	198m: 41307, 23604, 16522	330m: 68845, 39340, 27538
tungsten	99m: 20653, 11802, 8261	198m: 41307, 23604, 16522	330m: 68845, 39340, 27538
voltmeter	99m: 20653, 11802, 8261	198m: 41307, 23604, 16522	330m: 68845, 39340, 27538
ampere	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
oscillation	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076
waveform	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076

Appendix E

Estimates of ER Input, Reading Time for Repeated Encounters of Math Words

SVL: Math	**6 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**12 Encounters Input Size (words): 100, 260, 350wpm (hours)**	**20 Encounters Input Size (words): 100, 260, 350wpm (hours)**
width	1190286: 198, 76, 55	2380572: 396, 152, 113	3967620: 661, 254, 188
intercept	1201515: 200, 77, 55	2403030: 400, 154, 114	4005051: 667, 256, 190
verify	1224621: 204, 78, 56	2449242: 408, 157, 116	4082071: 680, 261, 194
horizontal	1286471: 214, 82, 59	2572942: 428, 164, 122	4288236: 714, 274, 204
acute	1286471: 214, 82, 59	2572942: 428, 164, 122	4288236: 714, 274, 204
radius	1768897: 294, 113, 81	3537795: 589, 226, 168	5896324: 982, 377, 280
equation	1768897: 294, 113, 81	3537795: 589, 226, 168	5896324: 982, 377, 280
quadratic	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
isosceles	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
integer	63.6m: 10613, 4082, 2948	127m: 21226, 8164, 6064	212m: 35377, 13606, 10107
cosine	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
quadrilateral	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
gradient	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m: 70755, 27213, 20215
exponential	127m: 21226, 8164, 5896	254m: 42453, 16328, 12129	424m:70755, 27213, 20215

Appendix F

Estimates of EV Input, Viewing Time for Repeated Encounters of Math Words

SVL: Physics	**6 Encounters Input Size (words): 80, 140, 200wpm (hours)**	**12 Encounters Input Size (words): 80, 140, 200wpm (hours)**	**20 Encounters Input Size (words): 80, 140, 200wpm (hours)**
probability	1223924: 254, 145, 101	2447848: 509, 291, 203	4079747: 849, 485, 339
domain	1348814: 281, 160, 112	2697629: 562, 321, 224	4496048: 936, 535, 374
integration	1406211: 292, 167, 117	2812421: 585, 334, 234	4687369: 976, 558, 390
axis	1406211: 292, 167, 117	2812421: 585, 334, 234	4687369: 976, 558, 390
translation	1586206: 330, 188, 132	3172411: 660, 377, 264	5287352: 1101, 629, 440
cone	1694664: 353, 201, 141	3389328: 706, 403, 282	5648881: 1176, 672, 470
consecutive	1724137: 359, 205, 143	3448273: 718, 410, 287	5747122: 1197, 684, 478
ratio	1819043: 378, 216, 151	3638086: 757, 433, 303	6063477: 1263, 721, 505
obtuse	33m: 6884, 3934, 2753	66m: 13769, 7868, 5507	110m: 22948, 13113, 9179
semicircle	39m: 8261, 4720, 3304	79m: 16522, 9441, 6609	132m: 27538, 15736, 11015
midpoint	39m: 8261, 4720, 3304	79m: 16522, 9441, 6609	132m: 27538, 15736, 11015
geometrical	49m: 10326, 5901, 4130	99m: 20653, 11802, 8261	165m: 34422, 19670, 13769
vertex	99m: 20653, 11802, 8261	198m: 41307, 23604, 16522	330m: 68845, 39340, 27538
quadratic	198m: 41307, 23604, 16522	396m: 82614, 47208, 33045	660m: 137691, 78680, 55076