Supporting Information for: Uchihara, T., Webb, S., & Yanagisawa, A. The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. Article accepted in *Language Learning* on 12 January 2019.

Appendix S1: Sensitivity Analysis

In order to examine the extent to which the current results were influenced by the presence of unpublished studies, we ran another analysis following the same procedure we adopted in the original analysis—that is, calculation of the mean effect size and moderator analysis—based on effect sizes from published studies (k = 32), with those from unpublished studies (k = 13) excluded.

To examine the overall relationship between frequency of encounters and vocabulary learning, the mean effect size was calculated, r = .37, 95% CI [.29, .45], p < .001 (see Figure 1). The magnitude of this effect size was slightly larger than that of the mean effect size found in the original analysis, r = .34, 95% CI [.27, .40], p < .001, but both were considered medium effects according to Plonsky and Oswald's (2014) criteria.

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Webb & Chang (2015)	-0.030	-0.280	0.224	-0.229	0.819
Daskalovska (2014a)	0.060	-0.419	0.513	0.233	0.816
Brown (1993)	0.063	-0.152	0.272	0.571	0.568
Daskalovska (2014b) sample 3	0.110	-0.254	0.447	0.584	0.559
Neuman & Koskinen (1992)	0.120	-0.239	0.450	0.649	0.516
Van Zeeland & Schmitt (2013) sample 2	0.204	-0.489	0.739	0.546	0.585
Zahar et al. (2001) sample 3	0.210	-0.138	0.512	1.187	0.235
Zahar et al. (2001) sample 2	0.220	-0.175	0.554	1.096	0.273
Yang & Sun (2013)	0.221	-0.024	0.441	1.769	0.077
Zahar et al. (2001) sample 4	0.240	-0.119	0.543	1.318	0.187
Daskalovska (2016) sample 1	0.250	-0.102	0.546	1.399	0.162
Van Zeeland & Schmitt (2013) sample 1	0.285	-0.181	0.646	1.206	0.228
Daskalovska (2014b) sample 1	0.290	-0.072	0.584	1.580	0.114
Ellis (1995) sample 1	0.290	-0.101	0.604	1.463	0.144
Daskalovska (2014b) sample 2	0.300	-0.054	0.587	1.667	0.096
Ellis (1995) sample 2	0.310	-0.107	0.634	1.469	0.142
Hatami (2017) sample 1	0.337	0.024	0.590	2.104	0.035
Daskalovska (2016) sample 2	0.340	-0.023	0.624	1.840	0.066
Vidal (2003)	0.343	0.171	0.494	3.800	0.000
Zahar et al. (2001) sample 1	0.400	-0.026	0.703	1.847	0.065
Pavia et al. (in press) sample 2	0.424	0.003	0.717	1.973	0.049
Tekmen & Daloglu (2006) sample 1	0.477	0.160	0.705	2.843	0.004
Vidal (2011) sample 2	0.488	0.301	0.639	4.681	0.000
Horst et al. (1998)	0.490	0.182	0.710	2.985	0.003
Hatami (2017) sample 2	0.497	0.256	0.679	3.775	0.000
Reynolds (2018) sample 1	0.511	0.327	0.658	4.918	0.000
Tekmen & Daloglu (2006) sample 3	0.539	0.246	0.742	3.356	0.001
Pavia et al. (in press) sample 1	0.545	0.045	0.827	2.117	0.034
Tekmen & Daloglu (2006) sample 2	0.563	0.267	0.762	3.432	0.001
Reynolds (2018) sample 2	0.633	0.488	0.744	6.882	0.000
Wenzhong & Feng (2009)	0.634	0.366	0.805	4.029	0.000
Vidal (2011) sample 1	0.687	0.575	0.774	8.793	0.000
	0.370	0.289	0.446	8.404	0.000

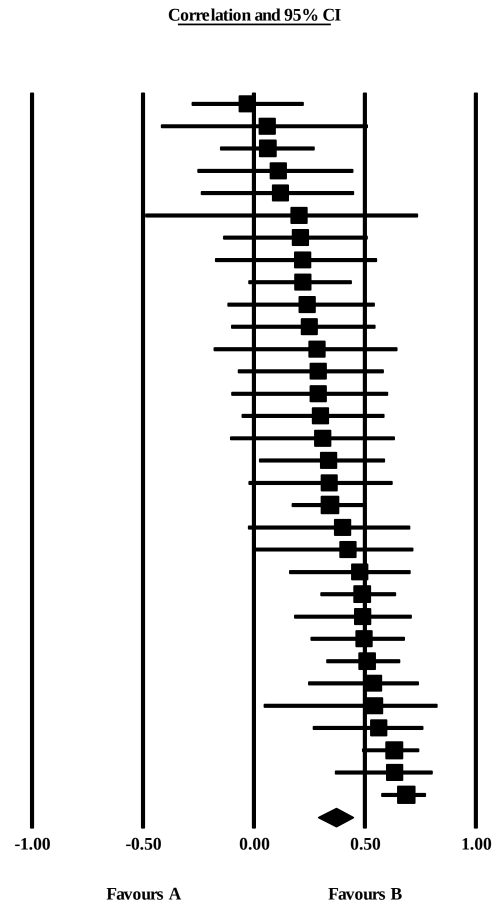

Figure 1 Overall average correlation between frequency of encounters and learning gains (indicated by a diamond) and correlations with confidence intervals for each study.

To explore whether and to what extent 10 variables moderate the relationship between frequency of encounters and vocabulary learning, a series of moderator analyses were conducted. The results of categorical variables and continuous variables are presented in Tables 1 and 2, respectively. Due to the decrease in the number of samples (i.e., effect sizes) and resulting decrease in statistical power, some moderator analyses which were found to be significant in the original analysis did not reach statistical significance ($\alpha = .05$). However, the direction of the relationship between moderators and frequency effects as well as the size of moderator effects were overall similar to the original results, except for moderators of age and basic vocabulary. As for the age variable, a relatively large effect was found for studies using primary school students ($r = .47$) compared to the effect found in the original analysis ($r = .20$). Yet, this result was based on two effect sizes ($k = 2$) from one study (Pavia, Webb, & Faez, in press); thus, the reliability of the effect is questionable. Regarding basic vocabulary, the regression and correlation coefficients (b = –0.0026, r = –.19) were smaller than the original analysis (b = –0.0046, r = –.43). However, considering the substantial decrease in the number of samples for this variable ($k = 22 \geq 14$) as well as the finding that the direction of the relationship was the same (i.e., negative correlation), this difference was considered minor.

In summary, the sensitivity analysis showed that the mean effect size of the relationship between frequency of encounters and vocabulary learning was similar to the original analysis. It was also found that the results of the moderator analysis were overall similar in terms of the direction as well as size of the moderating effects on the frequency-learning relationship. Therefore, we confirmed that there was little concern about the confounding influence of unpublished work on the current meta-analysis results.

Table 1 Results of analysis of categorical moderator variables

Variable	*k*	*r*	95% CI	*p*	*Q*(*df*)	*p*
<u>Learner</u>						
Age					9.85(31)	.001
Primary	2	.47	[.16, .69]	.001		
Secondary school	10	.20	[.09, .31]	.001		
University	20	.42	[.32, .51]	.001		
<u>Treatment</u>						
Spacing					2.00(31)	.150
Spaced learning	7	.26	[.06, .44]	.010		
Massed learning	25	.40	[.33, .48]	.001		
Mode of input					16.69(31)	.001
Reading	7	.49	[.38, .58]	.001		
Listening	10	.41	[.26, .54]	.001		
Reading while listening	13	.30	[.18, .41]	.001		
Viewing	2	.07	[−.10, .25]	.410		
Visual aid					2.87(31)	.090
Yes	6	.24	[.05, .41]	.010		
No	26	.40	[.32, .47]	.001		
Engagement					7.92(31)	.001
Yes	7	.21	[.07, .34]	.001		
No	25	.42	[.35, .50]	.001		
<u>Methodology</u>						
Nonword use					3.90(31)	.040
Yes	6	.50	[.36, .62]	.001		
No	26	.34	[.25, .42]	.001		
Comprehension test announcement					3.40(31)	.060
Yes	6	.51	[.33, .66]	.001		
No	26	.33	[.25, .40]	.001		
Test format					2.60(37)	.270
Recognition	25	.31	[.22, .39]	.001		
Recall	9	.42	[.25, .57]	.001		
Vocabulary knowledge scale	4	.45	[.22, .64]	.001		

Table 2 Results of weighted correlation analysis of continuous moderator variables

Variable	*k*	*r* (*rho*)	95% CI	*p*
Basic vocabulary	11	–.19 (–.20)	[−.91, .52]	.37
Range in encounters	24	–.50 (–.60)	[−.88, –.12]	.00

Appendix S2: Coding Scheme

Variables	Values				
Identification					
Author					
Title					
Year of publication					
Type of publication	Journal article	Dissertation	Book/book chapter	Proceeding	Presentation
Learner					
Age	Primary school	Secondary school	University		
Basic vocabulary					
Treatment					
Spacing	Spaced	Massed			
Mode of input	Reading	Listening	Reading while listening	Viewing	Other
Visual Aid	Yes	No			
Engagement	Yes	No			
Range in encounters					
Methodology					
Nonword use	Yes	No			
Comprehension test announcement	Yes	No			
Test format	Recognition	Recall	Vocabulary knowledge scale		
Sample sizes (*N*)					
Effect sizes (*r*)					

Note. Variables without labelled values are continuous, non-categorical, or open-ended.

Appendix S3: Results of Publication Bias Measures and Funnel Plot

We had a concern about overestimation of the aggregated effect sizes based on the fact that studies reporting statistically significant findings or large effect sizes tend to be published (Lipsey & Wilson, 2001). To assess publication bias in our data, we first employed a funnel plot, which provided information regarding the relationship between measurement precision (i.e., standard error or function of sampling error) and the effect sizes in question (i.e., correlation coefficient). Effect size distribution in a well-balanced data set has a symmetric funnel shape (i.e., the more sampling errors, the more likely that effect sizes vary across studies). Conversely, an unbalanced data set has an asymmetric funnel shape, an indication of publication bias. In the funnel plot, precision values (i.e., the inverse of the standard error or 1/standard error) are displayed on the y-axis and effect sizes for each study are displayed on the x-axis. Large sample studies (or effect sizes) generally appear towards the top of the graph and cluster around the mean effect size, whereas smaller studies spread out on the bottom half of the graph. In the presence of publication bias, studies normally miss on either side of the mean near the bottom.

In the current data, the funnel plot in Figure 1 displays an approximately equal distribution of the effect sizes around the mean, which suggests the absence of publication bias. In order to further examine whether the observed overall effect was robust in our data, we conducted the following three measures: (a) fail-safe *N*, (b) Orwin's fail-safe *N*, and (c) trim-and-fill. First, the classic fail-safe *N* (Rosenthal, 1979) calculates how many missing studies with small effect sizes (effect size = 0) would be needed before the p value for the observed mean effect became nonsignificant. The result showed that we would have needed to retrieve 2,182 studies to nullify the effect ($\alpha = .05$), indicating that there would be less reason for concern about publication bias. Second, Orwin's fail-safe *N* (Orwin, 1983) concerns the number of missing studies needed to lower the mean effect down to a certain effect-size value instead of focusing on statistical significance. The result showed that we would have needed to retrieve 122 studies before the mean effect size became the size of little importance (i.e., $r = .10$). This finding also suggested that the effect of publication bias was not serious in this study. Finally, we computed the trim-and-fill analysis (Duval & Tweedie, 2000a, 2000b) for the random-effects model. It calculated the unbiased estimate of the mean effect size by accounting for the lack of asymmetry in the funnel plot. The result showed that there appeared to be a slight underestimate of the mean effect size, but the adjusted value and initial value were overall similar: $r = .408$, 95% CI [.343, .469] and $r = .335$, 95% CI [.268, .398], respectively—both of the correlation coefficients were considered medium effects according to the criteria established by Plonsky and Oswald

(2014). This indicated no serious effect of publication bias on the current result. Therefore, findings based on the visual inspection of the funnel plot and the three bias-detective measures suggested that the concern for publication bias was very low.

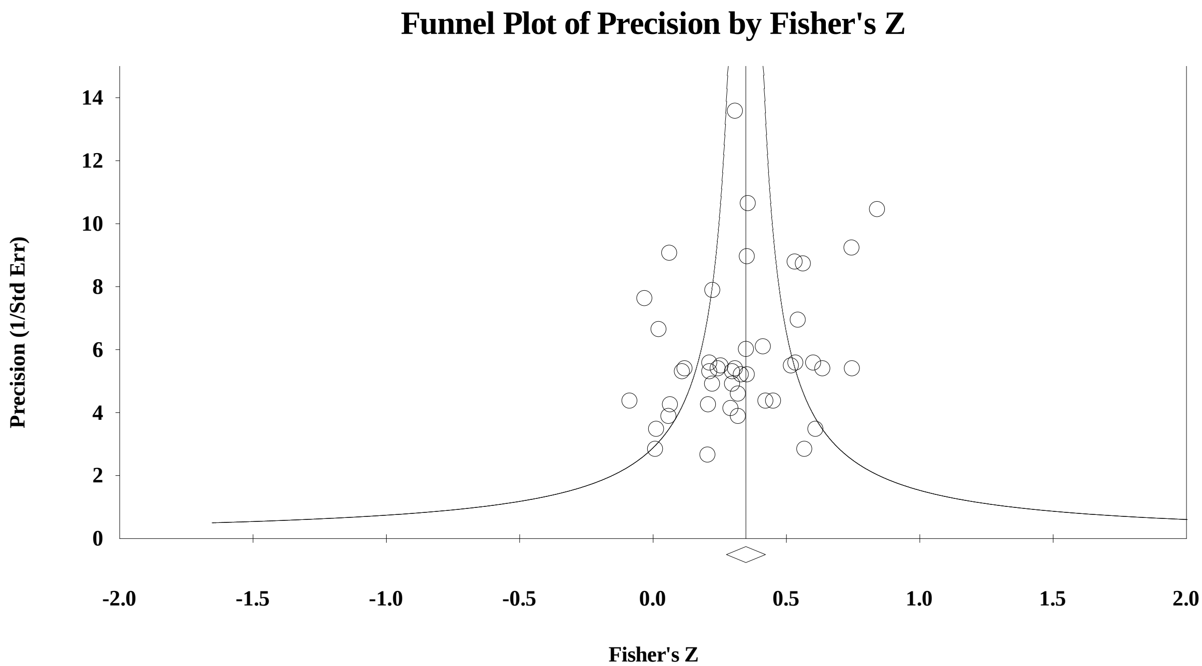

Figure 1 Funnel plot of effect sizes (*x*-axis) and standard error (*y*-axis)

References

Duval, S., & Tweedie, R. (2000a). A nonparametric trim and fill method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98. https://doi.org/10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics*, *8*, 157–159. https://doi.org/10.3102/10769986008002157

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Appendix S4: Relationship Between Frequency and Basic Vocabulary for 2,000 Frequency Level Mastered and Unmastered Samples

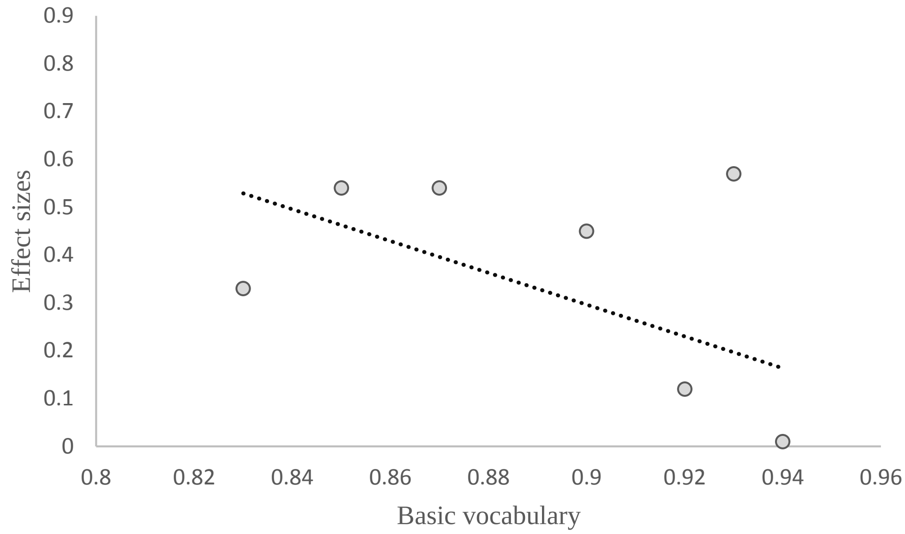

Figure 1 Relationship between basic vocabulary and effect sizes for mastered samples.

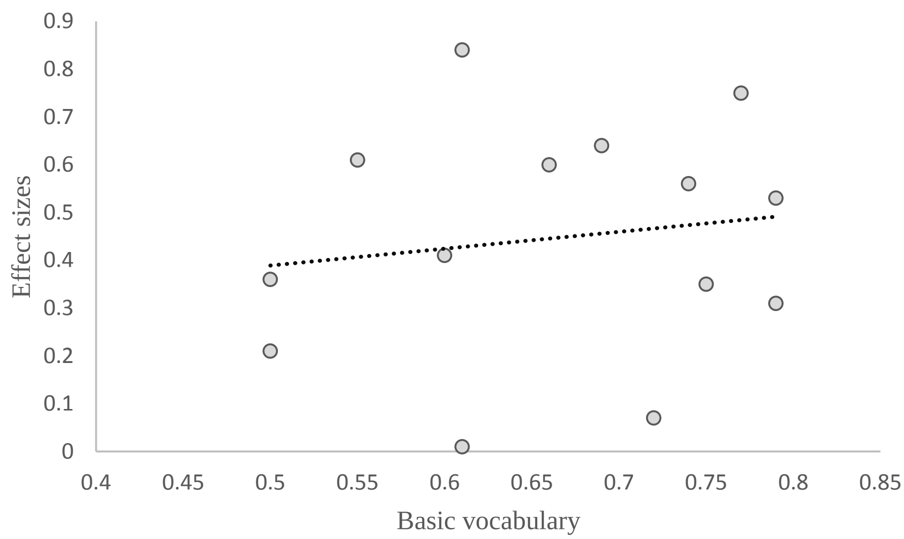

Figure 2 Relationship between basic vocabulary and effect sizes for unmastered samples.

Appendix S5: References of Included Studies

1. Alahirsh, H. (2014). *Exploring the effectiveness of extensive reading on incidental vocabulary acquisition by EFL learners: An experimental case study in a Libyan university.* (Unpublished doctoral dissertation). University of Nottingham, Nottingham, UK.

2. Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 263–286). Norwood, NJ: Ablex.

3. Daskalovska, N. (2010). Acquiring vocabulary through reading: A replication of the Golden Fleece study. In D. Köksal, I. H. Erten, E. Z. Topkaya, & A. Yavuz, A. (Eds.). The *6th International ELT Research Conference: Conference proceedings "Current trends in ESL research and language teaching*” (pp. 314-319). Selcuk-Izmir, Turkey: Tübitak.

4. Daskalovska, N. (2014a). Incidental vocabulary acquisition from reading an authentic text. *The Reading Matrix*, *14*, 201–216. Retrieved from http://eprints.ugd.edu.mk/11570/

5. Daskalovska, N. (2014b). Incidental vocabulary learning through reading. *Journal of Linguistics and Language Teaching*, *5*, 57–65. Retrieved from http://eprints.ugd.edu.mk/11106/

6. Daskalovska, N. (2016). Acquisition of three word knowledge aspects through reading. *Journal of Educational Research*, *109*, 68–80. https://doi.org/10.1080/00220671.2014.918530

7. Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, *16*, 409–441. https://doi.org/10.1093/applin/16.4.409

8. Feng, Y. (2017). *How does mode of input affect incidental vocabulary learning*. (Unpublished master’s thesis). University of Western Ontario, London, ON.

9. Ferris, D. (1988). *Reading and second language vocabulary acquisition*. (Unpublished manuscript), University of Southern California, Los Angeles, CA.

10. Hatami, S. (2017). The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge. *Reading in a Foreign Language*, *29*, 61–85.

11. Horst, M. (2000). *Text encounters of the frequent kind: Learning L2 vocabulary through reading*. (Unpublished doctoral dissertation). University of Wales, Swansea, UK.

12. Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.

13. Mohamed, A. (2009). *Investigating incidental vocabulary acquisition in ESL conversation classes*. (Unpublished master's thesis). University of North Texas, Denton, TX.

14. Neuman, S. B., & Koskinen, P. (1992). Captioned television as "comprehensible input": Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*, 95–106. https://doi.org/10.2307/747835

15. Pavia, N., Webb, S., & Faez, F. (in press). Incidental vocabulary learning from listening to L2 songs. *Studies in Second Language Acquisition*.

16. Pellicer-Sánchez, A., & Serrano, R. (2017, March). *Incidental L2 vocabulary learning from reading and reading-while-listening*. Paper presented at American Association of Applied Linguistics (AAAL), Portland, OR.

17. Reynolds, B. L. (2018). The effects of nonce words, frequency, contextual richness, and L2 vocabulary knowledge on the incidental acquisition of vocabulary through reading: More than a replication of Zahar et al. (2001) & Tekmen and Daloğlu (2006). *IRAL - International Review of Applied Linguistics in Language Teaching*, 1–28. Advance online publication https://doi.org/https://doi.org/10.1515/iral-2015-0115

18. Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage,*

attitudes, and captions. (Unpublished doctoral dissertation). Victoria University, Wellington, NZ.

19. Tekmen, E. A. F., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, *39*, 220–243. https://doi.org/10.1111/j.1944-9720.2006.tb02263.x

20. van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624. https://doi.org/10.1016/j.system.2013.07.012

21. Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*, 56–89. https://doi.org/10.1093/applin/24.1.56

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

23. Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*, 667–686. https://doi.org/10.1177/1362168814559800

24. Wenzhong, Z., & Feng, X. (2009). Effects of word exposure frequency on incidental acquisition of English vocabulary by Chinese EFL learners. *CELEA Journal*, *32*, 3–11.

25. Yang, H. C., & Sun, Y. C. (2013). It is more than knowledge seeking: Examining the effects of OpenCourseWare lectures on vocabulary acquisition in English as a foreign language (EFL) context. *Computer Assisted Language Learning*, *26*, 1–20. https://doi.org/10.1080/09588221.2011.624523

26. Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*, 541–572. https://doi.org/10.3138/cmlr.57.4.541