



#JALT2019 • NAGOYA 11.1-11.4

Building Vocabulary and Agency With an Online App

J. Tyler Kinkade

Kanagawa University

Reference Data:

Kinkade, J. T. (2020). Building vocabulary and agency with an online app. In P. Clements, A. Krause, and R. Gentry (Eds.), *Teacher efficacy, learner agency*. Tokyo: JALT.
<https://doi.org/10.37546/JALTPCP2019-58>

The impact of individualized vocabulary learning using an online spaced-repetition software application on the receptive lexical knowledge and agency of CEFR A2-level Japanese university students ($N = 17$) was investigated in this repeated-measures pilot study. The participants used the application to study vocabulary from one of five bands of the New General Service List (NGSL) based on their performance on the NGSL Test. A mixed-effects model indicated that learning vocabulary on the app predicted an increase on a delayed posttest at the end of the academic year—a large effect size. However, the effect of spaced review was not confirmed. All participants believed the app improved their English ability and the majority planned to continue using it independently in the future. The results suggest individualized online vocabulary study can contribute to vocabulary acquisition and promote learner agency.

この反復測定パイロット研究では、CEFRのA2レベルの日本人大学生 ($N=17$) の受容的な語彙知識と学習者主体に関して、オンラインの間隔反復ソフトウェアアプリを使用した個別の語彙学習の影響を調査した。参加者は、New General Service List (NGSL) テストのスコアによって、NGSLの5つのバンドの1つを選び、アプリを使用して、語彙を学習した。混合効果モデルでは、アプリで語彙を学習すると学期末の遅延ポストテストで成績向上(効果量大)が予測されることが示された。但し、間隔反復学習の効果は確認されなかった。すべての参加者は、アプリが英語能力を向上させ、大半が将来にアプリを単独で使用し続けるつもりだと考えた。その結果は、個別のオンライン語彙学習が語彙の習得に貢献し、学習者主体を促進できることを示唆している。

For English learners to acquire the approximately 3,000 most frequently used words needed for basic communication, Nation (2013) recommended spaced, bilingual flashcard study. Spaced-repetition software can make this flashcard practice even

more efficient (Godwin-Jones, 2010), and Memrise¹ (n.d.) is one such free internet/smartphone software application which permits teachers to monitor student progress. In Memrise, learners are first presented with vocabulary words or phrases with matching translations or definitions. These new words are then scheduled for spaced-repetition review. Notification messages, point awards, and dynamic graphics depicting learners' progress encourage the learners to continue learning and reviewing. Walker (2016) found that Memrise fostered acquisition and self-regulated learning of Latin vocabulary. However, it appears that published research has not measured its effect on receptive high-frequency lexical knowledge in repeated-measures designs using corpus-based lists, such as the New General Service List (Browne, Culligan, & Phillips, 2013), and corpus-based tests. The aim of the present paper is to contribute toward filling this gap by reporting on a repeated-measures pilot study examining the impact of individualized vocabulary learning with Memrise on receptive high-frequency vocabulary knowledge and learner agency.

New General Service List

The New General Service List (NGSL, Browne, Culligan, & Phillips, 2013) was created with the aim of compiling a list of the most useful words for English language learners. Its 2,801 headwords and their inflected forms provide 92% coverage of a 273 million-word subsection of the Cambridge English Corpus. In order to identify which specific subsections within the NGSL individual learners should study, Stoeckel and Bennett (2015) created the monolingual NGSL Test (NGSLT) based on the same specifications as Nation and Beglar's (2007) Vocabulary Size Test. The NGSLT is comprised of five 20-item levels corresponding to five 560-word frequency bands of the NGSL. The authors recommended learners focus on identifying and studying unknown words within the highest-frequency band for which they scored less than 80%. However, research published to date has not examined the effectiveness of this pedagogical approach.

Spaced Repetition

The advantage of delayed or spaced repetition over blocked learning (i.e., cramming) in acquiring knowledge and skills, such as vocabulary (Kornell, 2009), math skills (Rohrer, Dedrick, & Stershic, 2015), or art theory skills (Kornell & Bjork, 2008), has been well documented across disciplines. In addition, expanded rehearsal (i.e., gradually expanding the spacing over time) appears to have more enduring effects in vocabulary acquisition as long as learners receive immediate correction for errors (Pashler, Zarow, & Triplett, 2003).

Although providing immediate feedback for individualized vocabulary lists with expanded rehearsal would be difficult for teachers to implement in the classroom, spaced-repetition software (SRS) is well-suited to the task and its effectiveness in vocabulary learning has been investigated in a few studies. Hirschel and Fritz (2013) found that using online SRS to study vocabulary resulted in significant improvements in vocabulary acquisition compared to a control group. In addition, Bower and Rutson-Griffiths (2016) found a small correlation ($r = .34$) between TOEIC score gains and online SRS use to study TOEIC vocabulary.

Nakata (2011) set forth 17 criteria for evaluating SRS, including whether the software tests both receptive and productive recall and recognition, permits learner control of the number of words studied, adapts sequencing on the learner's responses, and gradually expands the intervals between tests (i.e., expanded rehearsal). Although it was not included in his study, Memrise meets all of his criteria except the criterion of having a feature testing generative use in different contexts—which none of the nine SRS programs he evaluated met either.

Most published research on Memrise focuses on students' perceptions of the app and its usage (e.g., Kent & Sherman, 2013; Ono, 2017). However, Bakla and Çekiç (2017), found that a group of learners who collaboratively created and studied word lists on Memrise had significantly higher self-reported lexical knowledge compared to a group that used only worksheets. Using paired *t* tests, Walker (2016) measured a larger effect in Memrise study of Latin words ($d = 1.2$) compared to traditional study methods ($d = .6$). However, both of these research studies made the unlikely assumption that each participant used Memrise to the same extent and in the same way, neglected individual differences, and did not quantify how much study on Memrise was needed to increase lexical knowledge. In fact, Memrise users can choose to only initially learn a word without continuing to use expanded spaced review, and it is possible for teachers and researchers to track such usage.

Promoting Learner Agency

According to Bandura (2001), agency involves setting, implementing, monitoring, and adjusting plans to achieve chosen goals. Thus, learner agency requires not only the autonomy to choose one's own goals, but also the ability to perceive the effectiveness of one learning strategy over another. However, Kornell and Bjork (2007) found that university students typically did not know effective learning strategies and the vast majority even chose ineffective strategies, such as cramming, over more effective ones, such as spaced learning, in spite of directly experienced evidence of differences in effectiveness. Thus, promoting learner agency is not simply a matter of giving students complete freedom over their learning. For this reason, Kornell and Bjork (2007) stressed the importance of teaching learning strategies and helping students to avoid subjective and misleading temporary results such as from the short-term effect of cramming. Based on Walker's (2016) finding that students generally recognized using Memrise as more effective than their previously preferred strategy of cramming, teaching and encouraging students to use SRS might be one way of implementing Kornell and Bjork's advice and thereby promote positive future expressions of learner agency.

In light of these observations, individualized vocabulary study with SRS offers a potentially powerful way to promote English language learners' agency. Rather than assigning learners identical vocabulary study lists, learners can identify for themselves useful high-frequency words that are unknown to them guided by tests such as the NGSLT. Such firsthand experience might help to change inaccurate deep-seated language learning beliefs and give learners the ability to choose spaced learning over cramming to attain future learning goals. Therefore, this study aimed to investigate the effectiveness of using Memrise in individualized vocabulary acquisition and in promoting learner agency by providing experience using expanded spaced review.

Hypotheses

The following hypotheses were proposed for this study:

1. Learning new words from a section of the New General Service List (NGSL) via the Memrise application results in increased receptive vocabulary knowledge for that section.
2. Expanded spaced review of the new words learnt from a section of the NGSL on the Memrise application results in increased receptive vocabulary knowledge for that section.
3. Most participants will recognize the value of using Memrise for independent vocabulary study after using it to study the NGSL.

Method

Participants

The participants were economics students ($N = 17$) in a required, two-semester, second-year integrated English skills course at a private university in the Kantō region. Their combined listening and reading TOEIC scores ranged from 295 to 460 ($M = 370$, $SD = 40$), which corresponds with an elementary (A2) level on the Common European Framework of Reference (Educational Testing Service, 2016). All students enrolled in the course completed the same curriculum and were evaluated without regard to their choice to voluntarily participate in the present study. Institutional approval and participants' informed consent were obtained.

Instruments

The instruments for the study were pre- and delayed posttreatment vocabulary tests, Memrise usage reports, and a posttreatment survey. The testing and treatment schedule is summarized in Table 1. In order to prevent test item familiarity effects, maximize lesson time, and reduce learner fatigue, the monolingual NGSLT was split into two 50-item tests. The odd-numbered items for each of the five levels of the NGSLT were administered as a pretest (hereafter referred to as NGSLT50A) with a 30-minute time limit in the second week of the course. The even-numbered items of the NGSLT were administered as a delayed posttest (hereafter NGSLT50B) in the 14th week of the second semester (i.e., 3 weeks after the treatment and 37 weeks after the pretest). Given the limited pool of NGSLT items to draw from, a delayed posttest research design was selected over an immediate posttest as the enduring vocabulary knowledge measured by a delayed posttest is arguably more valuable than the potentially short-lived knowledge measured by an immediate posttest.

The total number of new initially learned words and reviewed words on Memrise were recorded for each participant at the time of each learning target and the delayed posttest. In Memrise, a new initially learned word is one that the learner has both correctly spelled and matched to an equivalent translation or definition (in this study, an equivalent Japanese word) six times in quick succession, and a reviewed word is one that the learner has continued to correctly identify or spell in accordance with the application's expanded rehearsal schedule. Although the app prompts learners to review when words are due for rehearsal, learners may choose not to review at that time or to only learn new words without reviewing. Thus, once a new word has been initially learned it will always be included in the total of learned words even if it is never reviewed again; however, the total number of reviewed words will decrease whenever a learner fails to review.

Table 1. Testing and Treatment Schedule

Week	Test or treatment learning target
2	Pretest: NGSLT50A
3	Begin Memrise vocabulary study
4	No class (holiday)
5	Learning target: 10 reviewed words
15	Learning target: 110 reviewed words
17 – 24	No class (holiday)
26	Learning target: 110 reviewed words
36	Learning target: 210 reviewed words (final target)
37 – 38	No class (holiday)
39	Delayed posttest: NGSLT50B; survey

In addition to the above instruments, a bilingual Japanese-English survey administered after the delayed posttest included two 6-point Likert-response-format items directing the participants to rank from *strongly disagree* to *strongly agree* the degree to which the following statements aligned with their own views about Memrise use:

1. Studying vocabulary on Memrise helped improve my English ability.
2. In the future, I plan to study vocabulary on Memrise on my own.

Treatment

One week after the vocabulary pretest, participants were taught how to use Memrise and were assigned a specific section of the NGSLT to study on the app via bilingual English-Japanese quizzes outside of class.¹ The section assigned was the highest-frequency band in which the participant scored below 80% on the NGSLT pretest. The participants were directed to learn at least 10 new words each week and to review those words when prompted by Memrise so that the total number of reviewed words reported by Memrise matched the accumulative total (e.g., 30 words reviewed after three weeks of study). The final target was 210 reviewed words by the 12th week of the second semester (see Table 1).

Results

Hypotheses 1 and 2: Effects of Initial Learning and Review

Hypothesis 1 proposed that learning new words from a section of the NGSL via the Memrise application would result in increased receptive vocabulary knowledge for that section. The results support this hypothesis. Hypothesis 2 stated that expanded spaced review of the new words learnt from a section of the NGSL on the Memrise application would result in increased receptive vocabulary knowledge for that section. The results do not support this hypothesis.

Descriptive statistics for the independent (i.e., learned and reviewed words) and dependent (i.e., NGSLT scores) variables measured to test these hypotheses were calculated using the open-source statistical software R (R Foundation for Statistical Computing, 2019) and are reported in Table 2. Following LaFlair, Egbert, and Plonsky's (2015) recommendations for L2 research using small sample sizes, bias-corrected and accelerated (BCa) bootstrap confidence intervals (CI)² calculated using the Bootstrap R package (Canty & Ripley, 2019) are reported. The mean post NGSLT score for the studied section is higher and the CIs are fairly narrow and do not overlap. The mean number of words reviewed is lower than the words initially learned, but the CIs overlap somewhat and are wider for the number of reviewed words, which is expected as some participants chose not to review at all, as indicated in the Memrise usage reports.

Table 2. Descriptive Statistics

Variables	Pre		Post	
	<i>M (SD)</i>	BCa 95% CI	<i>M (SD)</i>	BCa 95% CI
NGSLT Studied Band (%)	51.2 (15.8)	[42.9, 57.6]	81.1 (15.8)	[72.4, 87.1]
Words Learned on App	0 (0)	-	194.5 (55.8)	[163.2, 216.6]
Words Reviewed on App	0 (0)	-	144.1 (86.5)	[95.7, 177.7]

Note. BCa = bias-corrected and accelerated bootstrap interval (10,000 replications)

Rasch model analysis performed with the Extended Rasch Modeling package (Mair, Hatzinger, & Maier, 2019) for R revealed person separation reliability indices of .80 and .75 for the NGSLT pretest and delayed posttest respectively, which are above Plonsky and Derrick's (2015) minimum suggested threshold of .74 for L2 research. This suggests

that both the pre- and delayed post-NGSLT were able to adequately detect differences between and within the participants.

To test Hypotheses 1 and 2, a mixed-effects model was applied using the Linear and Nonlinear Mixed-Effects Models package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019) in R. Mixed-effects modeling is an advanced form of regression analysis that can be used to account for both fixed and random effects, including repeated measure subject effects, and to determine whether treatment effects can be generalized to the population from which the participants were drawn (Cunnings & Finlayson, 2015). This makes it ideal for modeling complex designs such as in the present study, in which the number of learned and reviewed words varied by participant.

Based on Barr, Levy, Scheepers, and Tily's (2013) recommendations for mixed-effects confirmatory research of hypotheses, an a priori maximal model including random slopes for any repeated measures of interest was fit using maximum likelihood estimation.³ The model can be represented using Raudenbush and Bryk's (2002) notation as:

$$SCORE_{ti} = \beta_{00} + \beta_{11}(LEARNED_{1i}) + \beta_{12}(REVIEWED_{1i}) + \beta_{13}(LEARNED_{1i} \times REVIEWED_{1i}) + r_{0i} + r_{1i} \times LEARNED_{1i} + e_{ti}$$

In this model, the NGSLT score for the studied band of the NGSL at time *t* for participant *i* ($SCORE_{ti}$) is predicted by the intercept (β_{00}) plus the final number of initially learned words ($LEARNED_{1i}$), the final number of words which were not past due for expanded spaced rehearsal ($REVIEWED_{1i}$), and the interaction of these two variables multiplied respectively by the slope coefficients (i.e., effects) of learning new words (β_{11}), spaced review (β_{12}), and their interaction (β_{13}) plus the between-subjects errors of the intercept (r_{0i}) and slope (r_{1i}) and the within-subjects error (e_{ti}). In R, this model was coded as: `lme(SCORE ~ LEARNED*REVIEWED, random = ~LEARNED|PARTICIPANT, method = ML)`.

Plots were used to check the five assumptions for mixed-effects models (Crawley, 2007). First, a plot of fitted values versus standardized residuals appeared random. Second, response variables plotted against fitted values appeared linear. Third and fourth, quartile-quartile plots of the residuals and of the best linear estimated predictors of the random effects also appeared linear. Finally, plots of the random effects covariance matrix appeared random. Thus, there did not appear to be any violation of the five assumptions.

The model significantly predicted the outcome compared to a random-intercept null model, $\chi^2(5) = 27.00, p = .0001$ (the threshold for significance was set at $\alpha = .05$).

The model parameters are presented in Table 3. Following Larson-Hall and Plonsky’s (2015) recommendations for L2 research, effect sizes and CIs are included. Effect size was estimated as $r = \frac{t^2}{(t^2 + df)}$ (Rosnow & Rosenthal, 2000). Small, medium, and large effect size thresholds were $r = .25, .4,$ and $.6$ respectively based on Plonsky and Oswald’s (2014) recommendations for L2 research. The intercept parameter ($\beta_{00} = 51.5$ [43.5, 59.5]) and CI align closely with the mean pretest score and bootstrap CI. There was a significant ($p < .001$) effect for the number of words learned with a large effect size ($r = .77$), providing support for Hypothesis 1. The slope coefficient ($\beta_{11} = 0.21$ [0.11, 0.30]) indicates a NGSALT score increase of 21 percentage points is predicted for every 100 words initially learned. However, the effect of expanded spaced review of words ($r = .20$) was not statistically significant and below the threshold set for a small effect size, failing to provide support for Hypothesis 2.

Table 3. Mixed-Effects Model Parameters

Parameters	Fixed effects				Random effects	
	β	95% CI	t (df)	r	SD	95% CI
Intercept	51.5	[43.5, 59.5]	12.8 (16)	.95	14.0	[8.7, 22.6]
Words learned	0.21	[0.11, 0.30]	4.5 (14)	.77	0.06	[0.02, 0.14]
Words reviewed	0.17	[-0.28, 0.62]	0.76 (14)	.20	-	-
Learned \times reviewed	-0.001	[-0.003, 0.001]	-1.08 (14)	.28	-	-
Within-subjects SE	-	-	-	-	6.9	[2.3, 20.1]

Note. Effect size, r , is estimated as $\frac{t^2}{(t^2 + df)}$ (Rosnow & Rosenthal, 2000). SE = standard error.

Hypothesis 3: Recognizing Value of Memrise Study

Hypothesis 3 predicted that most participants will recognize the value of using Memrise for independent vocabulary study after using it to study the NGSL. The results support this hypothesis. To test Hypothesis 3, the percentage of responses for each option of the Likert-response-format items were calculated (see Table 4) and interpreted directly as further inferential statistical analysis is not justified for individual items (Brown, 2011). All of the participants believed that Memrise improved their English ability and 64% reported that they planned to continue using it independently in the future. These results provide support for the hypothesis that the participants will see value in using Memrise after using it for NGSL study.

Table 4. Memrise Survey Responses (%)

Item	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
Memrise helped improve my English ability.	0	0	0	35	41	24
I plan to study vocabulary on Memrise on my own.	0	18	18	35	29	0

Discussion

The results suggest that using Memrise to initially learn words from the NGSL leads to improvement of receptive vocabulary knowledge. An increase of 11 to 30 percentage points on the corresponding section of the NGSALT is predicted by the model for every 100 words learned. Importantly, this measured improvement was reflected in participants’ evaluation of Memrise’s impact on their English learning, in line with Walker’s (2016) results, and nearly two thirds reported that they planned to continue using it for independent study. Despite this and contrary to expectations, improvement in vocabulary knowledge did not result from expanded spaced review of those initially learned words. The changes in NGSALT section scores predicted by the model range from a decrease of 28 percentage points to an increase of 62 percentage points for every 100 words reviewed. Clearly, small sample size and wide variation in participants’ diligence in reviewing vocabulary—with some participants choosing not to review at all—is reflected in the results. The present study may not have had enough statistical power to determine whether review has an effect on receptive vocabulary knowledge as theorized and supported by previous research. In fact, all of these findings must be treated with caution due to the small sample size. Future research might lead to different conclusions.

Previous vocabulary acquisition research on Memrise did not measure the separate effects of initial learning versus spaced review, and thus, the current study contributes toward filling this gap. However, the results on the effectiveness of using Memrise to learn vocabulary are in line with Walker’s (2016) and Bakla and Çekiç’s (2017) more general findings. Although it is possible that Memrise’s effectiveness found in these previous studies was due to initial learning rather than subsequent spaced review, such an interpretation is challenged by studies which have found an advantage for spaced repetition over cramming (e.g., Kornell, 2009; Pashler, Zarow, & Triplett, 2003). However, initial learning is necessarily a step in both spaced repetition and cramming, so these



#JALT2019 • NAGOYA 11.1-11.4

Kinkade: *Building Vocabulary and Agency With an Online App*

studies did not measure the adequacy of initial learning alone in comparison to initial learning with subsequent review. Common sense and adages such as “practice makes perfect” lead to the expectation that any kind of review would be better than none, but it might also be that the gains in receptive lexical knowledge are small in comparison to the amount of review required. Another possible explanation is that the participants used other techniques to review and retain newly learned vocabulary by using it in communication. Clearly, more finely tuned research with larger sample sizes is required.

In regard to fostering learner agency through the promotion of accurate evaluation of learning strategies, the participants perceived that Memrise had a positive effect on their receptive vocabulary knowledge, which is in line with their increased test scores. Thus, Kornell and Bjork’s (2007) recommendation to teach students research-based learning strategies to counteract inaccurate learning beliefs appears to be supported by the results of the current study in which participants acknowledged the value of the spaced-repetition app. In this regard, giving students the opportunity to experience spaced-repetition study appears to contribute toward learner agency by helping them to identify effective vocabulary learning strategies. This claim could be further strengthened by the inclusion of survey items explicitly comparing Memrise with other vocabulary learning techniques; however, it could also be argued that evaluating the effectiveness of any particular technique requires participants to compare it with the techniques they have previously used in the seven years or more that they have been studying English. Future research could include survey items comparing Memrise’s effectiveness to other techniques and measuring learner agency in a repeated measures design to permit the use of inferential statistics and a clearer understanding of the app’s impact on learner agency. Delayed posttest surveys could also be used to determine whether participants continued to use the app independently.

Conclusion

This small pilot study was carried out in an attempt to better understand the impact of vocabulary learning and spaced review with the Memrise app on receptive lexical knowledge and learner agency. Overall, the results provide tentative support for the use of Memrise for learning vocabulary and fostering learner agency as well as Stoeckel and Bennett’s (2015) recommendations for individualized study based on NGSLT results. Although the effect of spaced review was not confirmed in this study, a large body of research supports the effectiveness of spaced repetition in other contexts. Additional research is needed to confirm these claims and generalize beyond the A2-level, Japanese university sophomore population from which the participants were sampled. Future

research should include a wider range of proficiency levels and a large enough sample size to model the random effects of spaced review and learning within subjects and across levels. The repeated measures design should also be expanded to include survey items measuring learners’ perception of the effectiveness of multiple vocabulary learning techniques as well as learners’ agency. Despite the limitations of the present study, Memrise’s potential for promoting independent vocabulary learning and agency certainly merit further investigation.

Notes

1. Teaching materials, including links to the Memrise vocabulary courses used in this study and directions for creating and monitoring student vocabulary courses, are posted at: www.researchgate.net/project/NGSL-NAWL-Memrise-Decks-Study
2. Bootstrapping involves taking a large number of random subset samples from a data set in order to improve the accuracy of confidence intervals (CI). The bias-corrected and accelerated method is one of several for calculating CI from bootstrapped samples. See LaFlair, Egbert, and Plonsky (2015) for a more detailed explanation.
3. Random participant and test level effects for the repeated measure of review are also possible. However, a data set larger than the one in this pilot study is required to fit a participant and test level hierarchical model with both random learning and reviewing effects.

Bio Data

Tyler Kinkade is an assistant professor in the Kanagawa University Faculty of Foreign Languages. In addition to vocabulary acquisition, he is interested in the development of diversity-related L2 sociolinguistic competence.

References

- Bakla, A., & Çekiç, A. (2017). Using an online vocabulary memorization tool versus traditional vocabulary exercises. *Ana Dili Eğitimi Dergisi [Journal of Mother Tongue Education]*, 5(4), 948-966. <https://doi.org/10.16916/aded.339241>
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1-26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>

Kinkade: *Building Vocabulary and Agency With an Online App*

- Bower, J. V., & Rutson-Griffiths, A. (2016). The relationship between the use of spaced repetition software with a TOEIC word list and TOEIC score gains. *Computer Assisted Language Learning*, 29(7), 1238-1248. <https://doi.org/10.1080/09588221.2016.1222444>
- Brown, J. D. (2011). What issues affect Likert-scale questionnaire formats? *Shiken*, 15(1). Retrieved from <http://teval.jalt.org>
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. Retrieved from <http://www.newgeneralservicelist.org>
- Canty, A., & Ripley, B. (2019). Bootstrap R (S-Plus) Functions (Version 1.3-22) [R software package]. Retrieved from <https://CRAN.R-project.org/package=boot>
- Crawley, M. J. (2007). *The R book*. New York, NY: Wiley.
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159-181). New York, NY: Routledge.
- Educational Testing Service. (2016). *Correlation table: TOEIC listening and reading test scores and the CEFR levels*. Retrieved from https://etswebsiteprod.cdn.prismic.io/etswebsiteprod%2Fddc2ab96-e7c6-41fb-bb3f-6b3f3d2bc825_toEIC+listening+and+reading+test++cefr+correlation+table.pdf
- Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: Approaches to second-language vocabulary learning. *Language, Learning & Technology*, 14(2), 4-11. <http://dx.doi.org/10.125/44208>
- Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, 41(3), 639-653. <https://doi.org/10.1016/j.system.2013.07.016>
- Kent, D., & Sherman, B. (2013). Pilot study for use of Memrise application by Korean junior college students studying EFL vocabulary in a blended learning context. *Yeongsang-yeong-eogyoyughaghoe [Society for Teaching English Through Media]*, 14(3), 169-192. <https://doi.org/10.16875/stem.2013.14.3.169>
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9), 1297-1317. <https://doi.org/10.1002/acp.1537>
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14(2), 219-224. <https://doi.org/10.3758/BF03194055>
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19(6), 585-592. <https://doi.org/10.1111/j.1467-9280.2008.02127.x>
- LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). *A practical guide to bootstrapping descriptive statistics, correlations, t-tests, and ANOVAs*. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159-181). New York: Routledge.
- Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(Supplement 1), 127-159. <https://doi.org/10.1111/lang.12115>
- Mair, P., Hatzinger, R., & Maier M. J. (2019). Extended Rasch Modeling (Version 1.0-0) [R software package]. Retrieved from <https://CRAN.R-project.org/package=eRm>
- Memrise. (n.d.). Memrise [Internet software application]. Retrieved from <https://decks.memrise.com>
- Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, 24(1) 17-38. <https://doi.org/10.1080/09588221.2010.520675>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13. Retrieved from http://jalt-publications.org/tlt/issues/2007-07_31.7
- Ono, T. (2017). *Vocabulary learning through computer-assisted language learning*. *Hitotsubashi Journal of Arts and Sciences*, 58(1), 67-72. <https://doi.org/10.15057/28989>
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1051-1057. <https://doi.org/10.1037/0278-7393.29.6.1051>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). Linear and Nonlinear Mixed Effects Models (Version 3.1-141) [R software package]. Retrieved from <https://CRAN.R-project.org/package=nlme>
- Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, 100(2), 538-553. <https://doi.org/10.1111/modl.12335>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>
- Quene, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413-425. <https://doi.org/10.1016/j.jml.2008.02.002>
- R Foundation for Statistical Computing. (2019). R (Version 3.6.1) [Computer software]. Retrieved from <https://www.R-project.org>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, 107(3), 900-908. <https://doi.org/10.1037/edu0000001>



#JALT2019 • NAGOYA 11.1-11.4

Kinkade: *Building Vocabulary and Agency With an Online App*

- Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes. *Zeitschrift Für Psychologie*, 217(1), 6-14. <https://doi.org/10.1027/0044-3409.217.1.6>
- Stoeckel, T., & Bennett, P. (2015). A test of the New General Service List. *Vocabulary Learning and Instruction*, 4(1), 1-8. <http://vli-journal.org/wp/wp-content/uploads/2015/10/vli.v04.1.2187-2759.pdf#page=6>
- Walker, L. (2016). The impact of using Memrise on student perceptions of learning Latin vocabulary and on long-term memory of words. *Journal of Classics Teaching*, 16(32), 14-20. <https://doi.org/10.1017/S2058631015000148>