Spacing Out Retrieval Practice in Adult Language Classes: A Randomized Classroom Trial

Mei-Lin Zhou[1], Aaron Kaplan[1], and Rosa Delgado[2]

[1]*Department of Educational Psychology, Northgate University;* [2]*Community Language Institute, Brookline Valley*

Abstract. Retrieval practice and spaced review each improve vocabulary learning in laboratory studies, but adult evening classes rarely have the time or control that those studies assume. We randomized 18 intermediate Spanish sections (N = 214 adults) to one of three review schedules: massed review in the final week, weekly spaced retrieval quizzes, or an expanding schedule of short quizzes at 1, 4 and 12 days after each unit. On a delayed vocabulary test eight weeks after the course, the expanding schedule retained 71% of items against 58% for weekly quizzes and 44% for massed review. Gains were largest for students who reported working more than 40 hours a week. We discuss what a part-time instructor can do with ten minutes of class time per session.

Keywords: spaced repetition, retrieval practice, adult education, second-language vocabulary, classroom trial

1. Introduction

Few findings in cognitive psychology are as well replicated as the spacing effect: material reviewed on several occasions separated in time is remembered longer than the same total study time packed into one session (Hartwell & Nunez, 2019). Retrieving an item from memory, rather than re-reading it, adds a second benefit. Taken together these results suggest a simple recipe for language teachers, yet adoption in adult education has been slow. Instructors of evening and weekend classes meet students once or twice a week for a limited time, students arrive tired, and attendance is uneven.

We asked whether a lightweight version of the recipe survives these conditions. Our questions were: (1) does an expanding schedule of brief retrieval quizzes outperform weekly quizzes and end-of-term massed review on delayed retention; (2) does the benefit depend on how much students can study outside class; and (3) do students find the expanding schedule acceptable.

2. Method

Participants. We recruited 18 sections of an intermediate Spanish course offered across six community institutes between January and May 2025. Of 261 enrolled adults, 214 (82%) completed the pre-test and at least 75% of sessions and are analysed here. Median age was 38 (range 19 to 71); 61% were women. Sections, not individuals, were randomly assigned to conditions (six sections each) using a seeded shuffle drawn by a person who did not know the instructors.

Materials. Each of eight units introduced 30 new words, 240 in all. Instructors used a common textbook and a common word list. Quizzes consisted of 10 cued-recall items presented on paper and marked in class; they took about ten minutes.

Conditions. In the *massed* condition students received no quizzes until a 90-minute review in week 15. In the *weekly* condition, each session began with a quiz drawn from the previous unit. In the *expanding* condition, items from each unit were quizzed 1, 4 and 12 days after the unit ended, with misses returned to the next quiz. Instructors received a printed schedule and a script; two sessions per section were observed to check fidelity.

Outcome. The primary outcome was the proportion of the 240 words recalled correctly on a cued-recall test given eight weeks after the last class. Testers were blind to condition. Students also completed a short questionnaire on hours worked per week and a five-point acceptability rating.

3. Results

Fidelity was high: observed sections delivered 94% of scheduled quizzes. Table 1 gives delayed-test performance by condition. A mixed-effects model with random intercepts for section showed a clear effect of condition (F(2, 15) = 9.8, p = .002). Pairwise comparisons favoured the expanding schedule over weekly quizzes (difference 13 points, 95% CI 5 to 21) and over massed review (27 points, 95% CI 18 to 36).

Condition	Sections	Students	Mean recall (%)	SD
Massed review	6	70	44	17
Weekly quizzes	6	71	58	15
Expanding schedule	6	73	71	13

Table 1. Delayed vocabulary test, eight weeks after the course.

Students who worked more than 40 hours a week (n = 58) gained most from the expanding schedule: they scored 24 points higher than their massed-review peers, compared with 17 points for students who worked fewer hours. The interaction was in the predicted direction but imprecise (p = .09). Acceptability was positive in all conditions; the expanding schedule averaged 4.1 of 5, and only 9% of students in that condition said the quizzes took too much class time.

3.1 Instructor and section effects

Because sections rather than students were randomized, we examined whether instructors drove the result. Each of the six instructors taught three sections, spread over at least two of the three conditions. Adding instructor as a random effect changed the condition estimates by less than one percentage point. The intraclass correlation at the section level was .11, which is typical for classroom data. Section size ranged from 9 to 16 students and was unrelated to condition (r = .04).

3.2 Item-level analysis

Words differed in difficulty. Cognates (for example *hospital*, *importante*) were recalled at 81% overall, whereas non-cognates were recalled at 49%. The expanding schedule helped both kinds, but the advantage over massed review was larger for non-cognates (31 points) than for cognates (19 points). Words introduced in the first two units, which had the longest delay before the final test, showed the largest differences across conditions: 38 points between expanding and massed review, against 17 points for words from the last two units.

Word type	Massed	Weekly	Expanding
Cognates (n = 72)	68	77	87
Non-cognates (n = 168)	34	49	65
Units 1-2 (n = 60)	31	52	69
Units 7-8 (n = 60)	58	63	75

Table 2. Mean percent recalled at the delayed test by word type and unit position.

3.3 Attrition and missing data

Of the 261 enrolled students, 47 were excluded because they missed the pre-test or attended fewer than 75% of sessions (massed: 21 of 91; weekly: 14 of 85; expanding: 12 of 85). We re-ran the primary analysis with multiple imputation under a missing-at-random assumption, including pre-test score and hours worked as predictors. The condition effect was nearly unchanged (expanding minus massed: 25 points, 95% CI 15 to 35). A sensitivity analysis assuming that excluded students in the massed condition would have scored 10 points lower than the imputation implied still left the expanding schedule ahead by at least 20 points.

4. Discussion

In ordinary adult classes, spacing short retrieval quizzes over two weeks after each unit produced better long-term retention than either weekly quizzes or massed review, with no added class time. The result matches laboratory findings and extends them to a setting where attendance and out-of-class study vary widely. We think the main practical message is that *when* an item is quizzed matters as much as *whether* it is quizzed.

Limitations. Randomizing by section means that instructor differences are partly confounded with condition, although the six instructors each taught sections in two conditions. Our test was cued recall of single words; we did not measure use of the words in speech or writing. Retention was measured at eight weeks, and longer follow-up is needed. Attrition was 18% and slightly higher in the massed condition.

4.1 What an instructor can do on Monday

We close with four suggestions that came out of our debrief with the six instructors. First, keep a simple card file or spreadsheet of units and their quiz dates; instructors who used a printed calendar delivered 97% of scheduled quizzes, against 88% for those who relied on memory. Second, put the quiz at the start of class, when students are arriving and the first ten minutes are otherwise lost. Third, return missed items to the next quiz rather than re-teaching them; students said this felt fairer than a repeated lecture. Fourth, tell students why: in the sections where the instructor explained the reason for the spacing, acceptability ratings were 0.4 points higher.

4.2 Costs

The expanding schedule required about 25 minutes of preparation per unit for a typical instructor, mostly spent selecting items and photocopying. Two instructors built quizzes from a shared spreadsheet and reduced this to under ten minutes. Grading took three to four minutes per quiz when students swapped papers. We estimate the net cost at roughly two hours of instructor time over a 16-week course, which institutes may reasonably regard as small relative to the retention gain.

5. Conclusion

A ten-minute expanding quiz schedule is a cheap and acceptable way to improve vocabulary retention in adult language classes. We provide the schedule and quiz templates to instructors on request.

Acknowledgements

We thank the instructors and students of the six institutes. This work was supported by a small grant from the Northgate University Teaching Fund.

References

Hartwell, J., & Nunez, P. (2019). Distributed practice in applied settings: a review of field studies. *Journal of Applied Learning Research, 12*(3), 201-224.

Ibarra, S., Lindqvist, K., & Moore, T. (2022). Retrieval practice and adult learners' working hours. *Adult Education Quarterly Review, 31*(1), 45-62.

Okonkwo, B. (2021). Brief quizzes in community language classes. *Language Teaching in Practice, 8*(2), 90-107.

Tran, H. V., & Weiss, A. (2018). Cued recall and delayed testing of second-language vocabulary. *Studies in Second Language Memory, 5*(4), 311-338.

Appendix A. Expanding schedule used in the trial

Unit ends	Quiz 1 (+1 day)	Quiz 2 (+4 days)	Quiz 3 (+12 days)
Unit 1 (week 2)	Week 3, session 1	Week 3, session 2	Week 5
Unit 2 (week 4)	Week 5, session 1	Week 5, session 2	Week 7
Unit 3 (week 6)	Week 7, session 1	Week 7, session 2	Week 9

Unit 4 (week 8)	Week 9, session 1	Week 9, session 2	Week 11

Table A1. Units 5 to 8 follow the same pattern. Sessions are twice weekly; quizzes use ten items from the target unit.

Appendix B. Questionnaire items

1. About how many hours a week do you work for pay? (0, 1-10, 11-20, 21-30, 31-40, more than 40)

2. How many hours a week do you study Spanish outside class? (0, 1-2, 3-4, 5 or more)

3. The review activities in this class were a good use of class time. (1 strongly disagree to 5 strongly agree)

4. The quizzes made me more confident about remembering new words. (1 to 5)

5. I would like my next language class to use the same review method. (1 to 5)