

Instructor: Sean Kang, Ph.D.

Email: sean.kang@dartmouth.edu

Class Meets: Weekdays 1:30-4:00pm @ 9B114

SKKU ISS3147

Myths and Mysteries of Human Learning and Memory

Testing and Spacing:
Keys to Effective Learning

30 Jun 2016

1

Purpose of Tests / Quizzes

- Traditionally, an assessment tool
- But testing does not merely measure the contents of memory
- Taking a test can serve as a learning opportunity, enhancing memory retention to a greater extent than additional studying...

the **testing effect**

(also referred to as the benefit of *retrieval practice*)

The Testing Effect

- Demonstrated with diverse study stimuli:
 - Word lists
 - Paired associates
 - General knowledge facts
 - Prose passages
 - Visuo-spatial info

ON THE ANALYSIS OF THE FACTOR OF RECALL
IN THE LEARNING PROCESS.

BY EDWINA E. ABBOTT(1909)

Spitzer (1939)

- 3,605 sixth-graders in Iowa
- Students read ~600-word article on the bamboo plant
- 25-item multiple-choice test (no feedback)
- Varied the retention interval and frequency of testing

Spitzer (1939)

Group	Time After Studying (Days)						
	0	1	7	14	21	28	63
1	T ₁	T ₂			T ₃		
2	T ₁		T ₂				T ₃
3		T ₁		T ₂			
4			T ₁		T ₂		
5				T ₁		T ₂	
6					T ₁		T ₂
7						T ₁	
8							T ₁

5

Spitzer (1939)

Group	Time After Studying (Days)						
	0	1	7	14	21	28	63
1	13.2						
2	13.2						
3		9.6					
4			7.9				
5				7.0			
6					6.5		
7						6.8	
8							6.4

6

Spitzer (1939)

Group	Time After Studying (Days)						
	0	1	7	14	21	28	63
1	13.2	13.1			12.2		
2	13.2		11.8				10.7
3		9.6		8.9			
4			7.9		8.2		
5				7.0		7.1	
6					6.5		7.1
7						6.8	
8							6.4

7

“A curious peculiarity of our memory is that things are impressed better by active than by passive repetition. I mean that in learning (by heart, for example), when we almost know the piece, it pays better to wait and recollect by an effort from within, than to look at the book again. If we recover the words in the former way, we shall probably know them the next time round; if in the latter way, we shall very likely need the book once more.”

(William James, 1890, *Principles of Psychology*)

“If you read a piece of text through twenty times, you will not learn it by heart so easily as if you read it ten times while attempting to recite from time to time and consulting the text when your memory fails.”

(Sir Francis Bacon, 1620, *Novum organum*)

“Exercise in repeatedly recalling a thing strengthens the memory.”

(Aristotle, 4th century B.C., *De Memoria et Reminiscentia*)

The Testing Effect

- Journal of Educational Psychology, 1989:
The “Testing” Phenomenon: Not Gone but Nearly Forgotten

John A. Glover
Teachers College and Burriss Laboratory School
Ball State University

- Dempster, F. N. (1992). Using tests to promote learning: A [neglected classroom resource](#). *Journal of Research & Development in Education*, 25, 213–217.
- Resurgence of interest in the testing effect in recent years

Test-Enhanced Learning

Taking Memory Tests Improves Long-Term Retention

Henry L. Roediger, III, and Jeffrey D. Karpicke

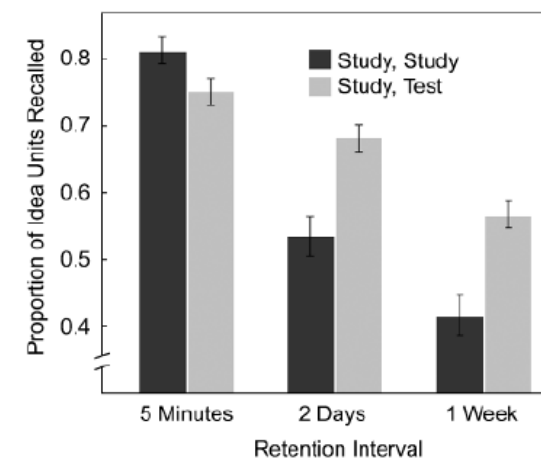
Washington University in St. Louis



10

Roediger & Karpicke (2006)

- Stimuli: 2 prose passages from TOEFL prep book (~260 words each)
- Learning condition (within-subjects):
 - Restudy (two 7-min periods of study) vs. Test (7-min period of study, followed by 7-min period of test)
- Retention interval (between-subjects):
 - 5 min, 2 days, or 1 week



11

12

Testing beyond words: Using tests to enhance visuospatial map learning

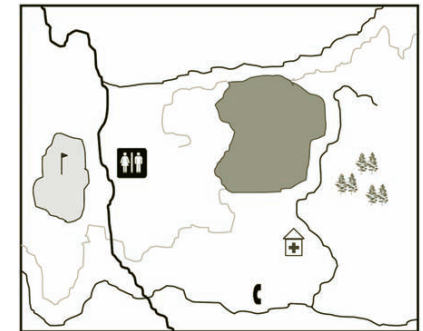
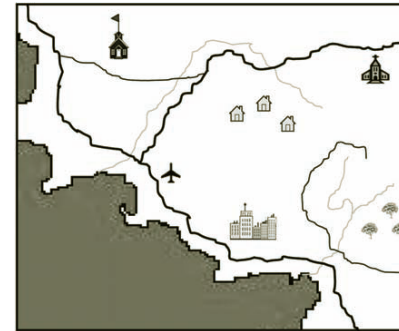
SHANA K. CARPENTER AND HAROLD PASHLER
University of California, San Diego, La Jolla, California



13

Does testing benefit memory for non-verbal materials?

- Past research has focused exclusively on verbal materials (or at least required verbal responses at test)



Carpenter & Pashler (2007)

14

Mean Proportion Correct on Final Test As a Function of Scoring Procedure, Counterbalancing Condition, and Test Condition

Scoring Procedure	Counterbalancing Condition								Total	
	1		2		3		4		M	SE
	M	SE	M	SE	M	SE	M	SE		
Test/Study										
LQ	.82	.04	.68	.06	.73	.05	.75	.06	.74	.03
SQ	.69	.04	.56	.05	.58	.05	.63	.05	.61	.02
LRP	.76	.05	.61	.06	.65	.05	.72	.06	.68	.03
SRP	.68	.04	.55	.06	.58	.05	.63	.06	.61	.03
Study										
LQ	.64	.06	.66	.07	.62	.06	.62	.07	.63	.03
SQ	.54	.05	.52	.07	.53	.06	.53	.07	.53	.03
LRP	.63	.05	.62	.07	.67	.06	.56	.07	.62	.03
SRP	.56	.05	.54	.07	.57	.06	.51	.07	.54	.03

Note—The maps were scored according to a liberal quadrant (LQ) procedure, a stringent quadrant (SQ) procedure, a liberal relative position (LRP) procedure, and a stringent relative position (SRP) procedure. All four scoring procedures yielded a significant benefit of test/study over study.

15

The Critical Importance of Retrieval for Learning

Jeffrey D. Karpicke^{1*} and Henry L. Roediger III²

15 FEBRUARY 2008 VOL 319 SCIENCE www.sciencemag.org

16

Roediger & Karpicke (2008)

17

- Stimuli: 40 Swahili-English word pairs
- Subjects studied and were tested on the Swahili words in alternating blocks

Condition	Study (S) or test (T) period and number of trials per period								Total number of trials
	1	2	3	4	5	6	7	8	
ST	S 40	T 40	S 40	T 40	S 40	T 40	S 40	T 40	320
S _N T	S 40	T 40	S _N 26.8	T 40	S _N 8.0	T 40	S _N 2.0	T 40	236.8
ST _N	S 40	T 40	S 40	T _N 27.9	S 40	T _N 11.8	S 40	T _N 3.3	243.0
S _N T _N	S 40	T 40	S _N 27.1	T _N 27.1	S _N 8.8	T _N 8.8	S _N 1.5	T _N 1.5	154.8

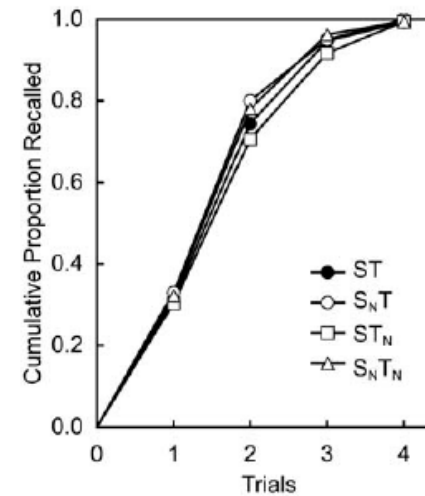


Fig. 1. Cumulative performance during the learning phase.

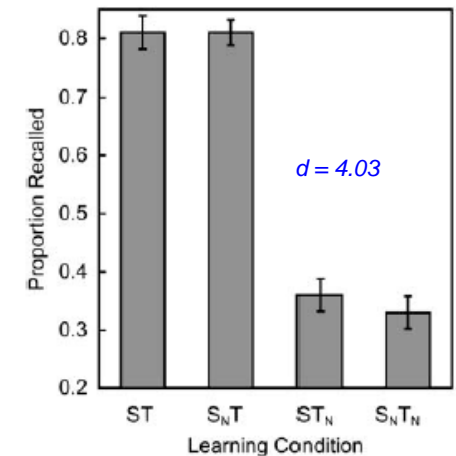


Fig. 2. Proportion recalled on the final test 1 week after learning. Error bars represent standard errors of the mean.

18

Testing effect: How does it work?

- Additional (focused) presentation of material
- Operations/processes engaged by an initial test are also engaged during the final test, resulting in positive transfer to same type of tests (i.e., practice effect)
- Retrieval itself is a potent memory modifier, with increasing retrieval demand/effort enhancing later retention

19

Test format and corrective feedback modify the effect of testing on long-term retention

Sean H. K. Kang, Kathleen B. McDermott, and Henry L. Roediger, III

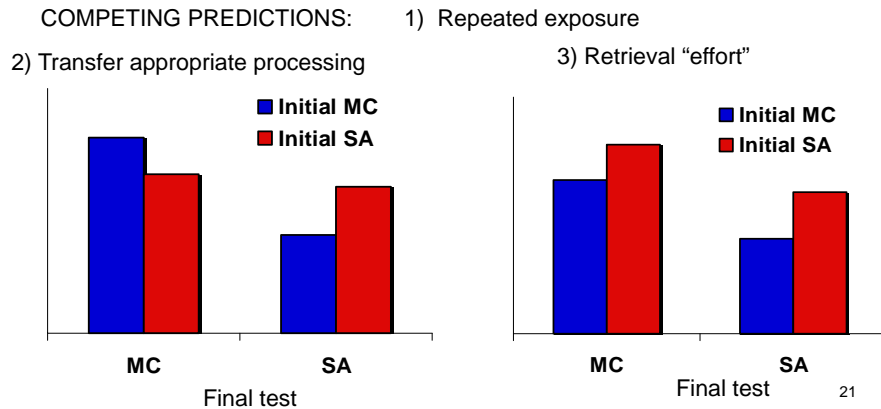
Washington University in St Louis, St Louis, MO, USA



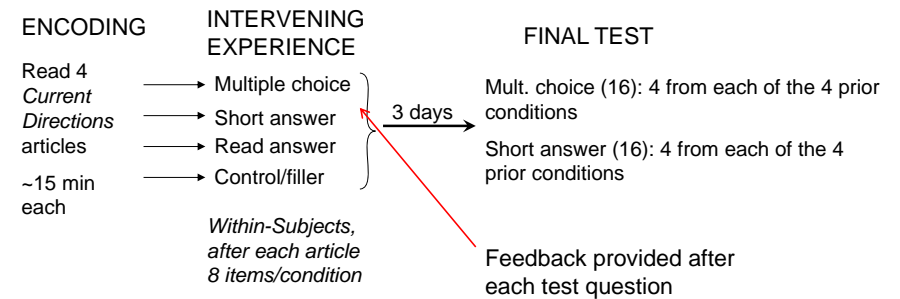
20

Does test format matter?

Initial test type -- Short Answer (SA), Multiple Choice (MC), Read Fact
 Final (criterial) test -- SA, MC
 (Corrective feedback given after each initial test question)



Procedure

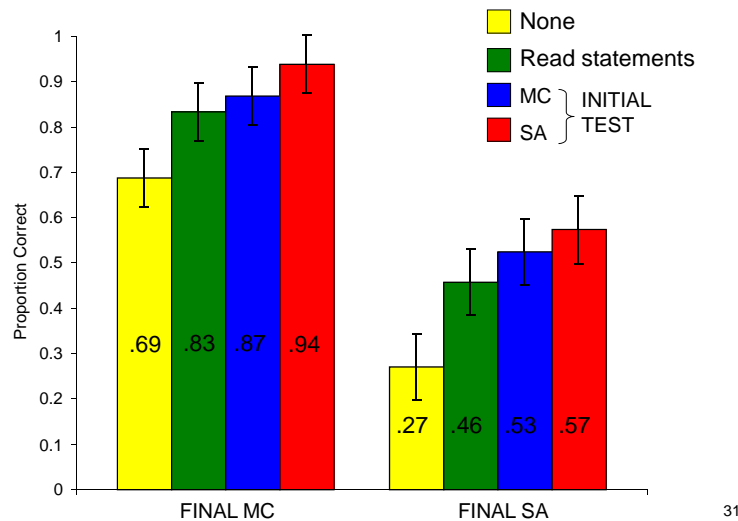


N=48

(Kang, McDermott, & Roediger, 2007)

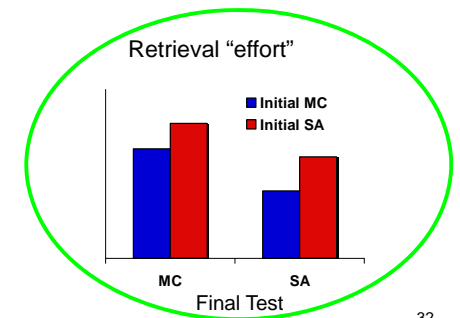
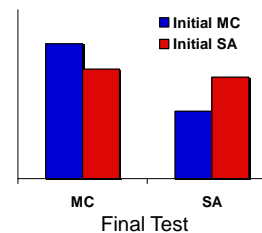
22

Testing enhanced later memory, and the enhancement was greater when the initial test format was short answer

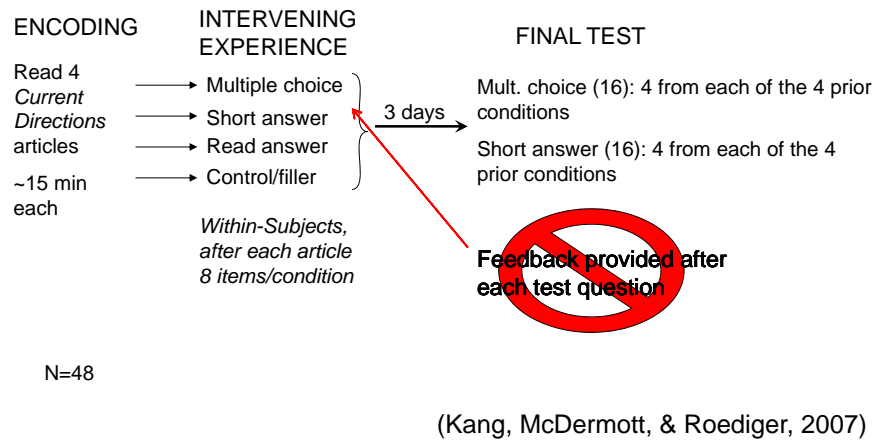


COMPETING PREDICTIONS:

Transfer appropriate processing

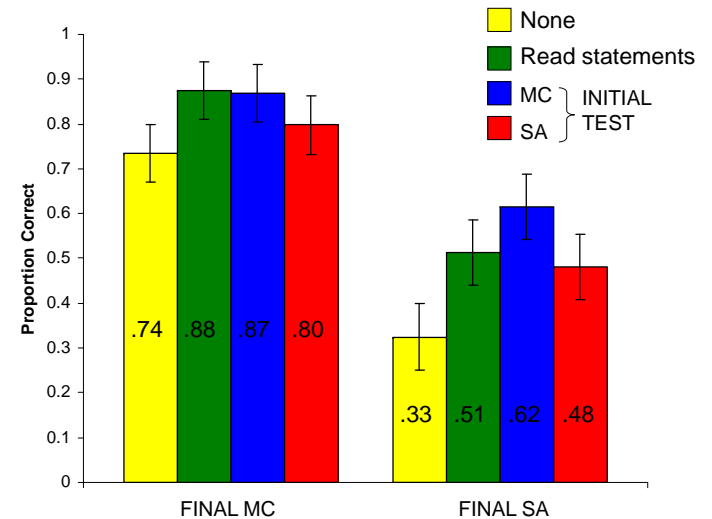


Does feedback matter?



33

Corrective feedback important, especially when initial test performance is not high



34

Don't just repeat after me...

Research question:

When learning to speak a foreign language, should you practise by retrieving from memory or by imitating what you hear?

(Kang, Gollan, & Pashler, 2013)



Tamar Gollan
UCSD
Psychiatry

Hal Pashler
UCSD
Psychology



Foreign Language (L2) Instruction

Informal survey of popular L2 instruction programmes revealed that *imitation* is a common pedagogical technique.

- E.g., Berlitz Guaranteed, The Pimsleur Method, Linguaphone, Rosetta Stone
- Immediately after hearing an L2 word/phrase spoken by a native speaker, the learner repeats it aloud.

But is imitation an effective method?

36

Retrieval Practice (aka the Testing Effect)

- Benefits of retrieval practice are well established:
 - Learning L1 translations of L2 words (e.g., Karpicke & Roediger, 2008)
 - Learning L2 orthography (Kang, 2010)

fight	斗	wood	木
open	开	door	门
car	车	cow	牛

- Would the advantage of retrieval practice extend to the learning of *spoken* vocabulary? (especially when compared against an active learning condition, i.e., imitation)

37

Kang, Gollan, and Pashler (2013)

Stimuli: 10 Hebrew nouns each from 4 categories

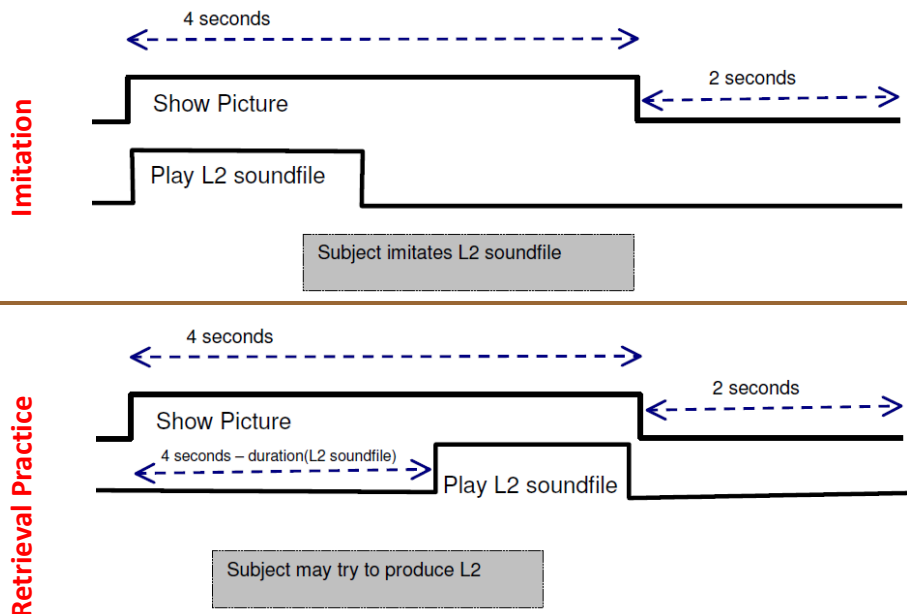
- Body parts (e.g., ear, hand)
- Eating/food (e.g., fork, bread)
- Animals (e.g., dog, elephant)
- Household objects (e.g., clock, chair)

Design:

- Items were blocked by category, and each category was trained in separate blocks.
- 2 training conditions (Retrieval Practice vs. Imitation), manipulated within-Ss.
- In each block, there were 3 cycles of training in the assigned condition.
- Final test (immediately after training on all 4 categories) assessed comprehension and production.

38

Procedure



Results – Training Performance

Production during final (3rd) cycle of practice

	<u>Proportion Correct</u>	
Retrieval Practice	.35	
	^	$d = 3.27$
Imitation	.90	

Pronunciation quality when correct (rated on 10-pt. scale)

	<u>Pronunciation Quality</u>	
Retrieval Practice	5.7	
	^	$d = 0.58$
Imitation	6.5	

43

Results ($N = 41$) – Final Test

Comprehension (accessing semantics when presented with phonology)

Subject heard soundclip of Hebrew word, and asked to select the meaning (array of 40 thumbnail pictures displayed on the screen—*new* exemplars of the previously presented objects).

	<u>Proportion Correct</u>	
Retrieval Practice	.63	$d = 0.31$
Imitation	.57	

44

Results ($N = 41$) – Final Test

Production (retrieving phonology when given semantics)

Subject presented with picture of object and asked to pronounce aloud the Hebrew word.

	<u>Proportion Correct</u>	
Retrieval Practice	.40	$d = 0.62$
Imitation	.27	

Pronunciation quality when correct (rated on 10-pt. scale)

	<u>Pronunciation Quality</u>
Retrieval Practice	5.6
Imitation	5.4

45

The pattern of results was replicated in a 2nd experiment which featured more training (6 cycles) and a 1-week delay before the final test was administered.

Conclusion

When learning to speak a foreign language, should you practise by retrieving from memory or by imitating what you hear?

(Kang, Gollan, & Pashler, 2013)

46

Any *negative* side effects of testing?

Multiple-choice and True/False tests expose students to wrong (plausible) answers.

Remmers and Remmers (1926) first raised the possibility that lures on a T/F test would cause a “negative suggestion effect”:

“it is obvious that the opportunity for forming wrong connections exists in approximately 50 percent of the test items... do true-false tests tend to leave a residue of false associations?”

Exposure to false statements on T/F and MC tests increased subjects validity ratings of these statements 1 week later (Toppino & Brochin, 1989; Toppino & Luipersbeck, 1993).

47

The Positive and Negative Consequences of Multiple-Choice Testing

Henry L. Roediger III
Washington University in St. Louis

Elizabeth J. Marsh
Duke University



Multiple-choice tests are commonly used in educational settings but with unknown effects on students' knowledge. The authors examined the consequences of taking a multiple-choice test on a later general knowledge test in which students were warned not to guess. A large positive testing effect was obtained: Prior testing of facts aided final cued-recall performance. However, prior testing also had negative consequences. Prior reading of a greater number of multiple-choice lures decreased the positive testing effect and increased production of multiple-choice lures as incorrect answers on the final test. Multiple-choice testing may inadvertently lead to the creation of false knowledge.



Experiment procedure:

1. Read/study material
2. Take MC test (answer every question, even if not sure)
3. Filler task (i.e., brainteaser puzzles)
4. Take cued-recall test (warned not to guess when unsure)

48

Positive Effects of Testing

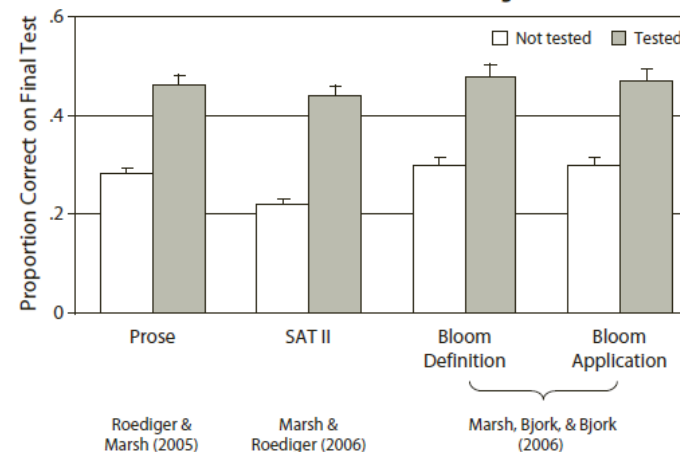


Figure 1. Proportion correct on the final general knowledge test as a function of whether or not concepts had been tested previously, for different types of materials.

49

Negative Effects of Testing

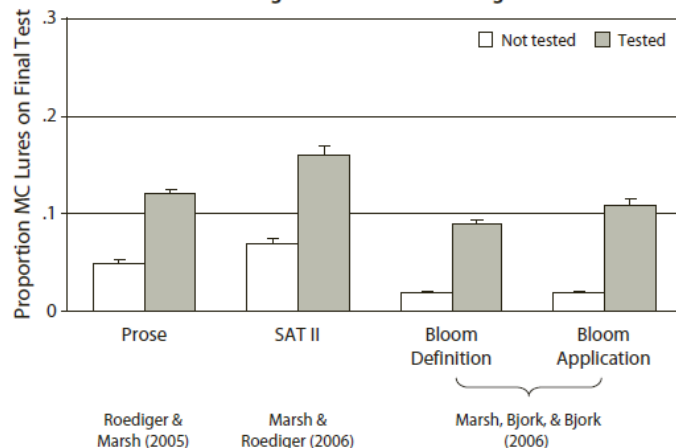


Figure 2. Intrusions of multiple-choice (MC) lures on the final general knowledge test as a function of whether or not concepts had been tested previously, for different types of materials.

Butler & Roediger (2008) showed that providing feedback after the MC test eliminated this negative effect of lures intruding on a later cued-recall test.

50

Summary

- Taking a test can be a potent learning event, often yielding better long-term retention than additional studying.
- Testing benefits learning of a diverse range of materials, both verbal and nonverbal.
- *Repeated* retrieval practice augments the benefit.
- The size of the testing effect is modulated by test format & feedback
 - Tests requiring effortful retrieval are more effective at enhancing retention, implicating retrieval as a causal mechanism
 - To maximize the benefit of testing, feedback should be provided when initial test performance is low

51

Summary (cont'd)

- Lures on T/F and MC tests can lead to students learning wrong information
 - Corrective feedback reduces this negative side effect

52

Summary

- Other (indirect) benefits of testing:
 - Helps identify gaps in knowledge
 - Provides feedback to instructors
 - Encourages students to study

The Atlantic



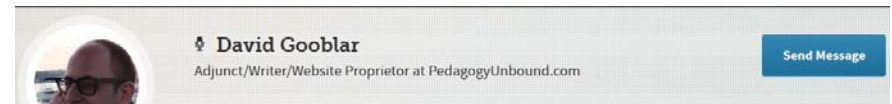
Print | Close

Students Should Be Tested More, Not Less

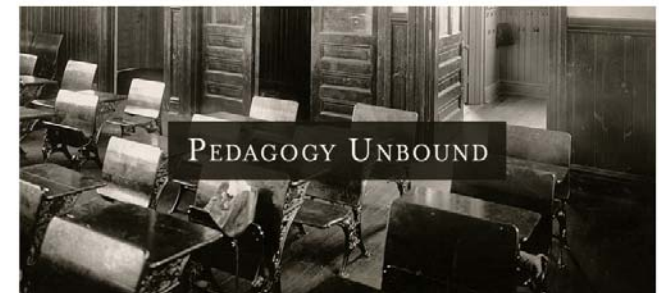
By Jessica Lahey



Bebeto Matthews/AP Images



We Should Give Students More Tests. Seriously.



February 5, 2014

I can't believe I'm typing these words, but I'm here today to convince you to give your students more tests.

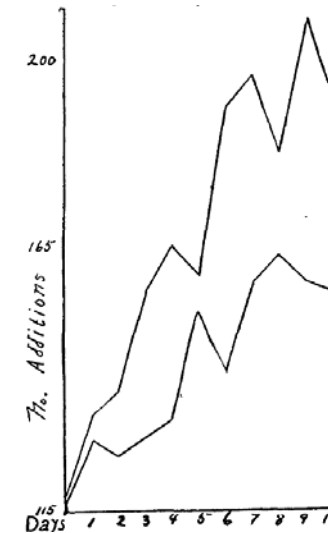
The Spacing Effect

- Reviews are more effective when distributed or spaced out, rather than massed (with total time equated)
- One of the most robust phenomena; observed with diverse range of materials / types of learning



- Ebbinghaus (1885):

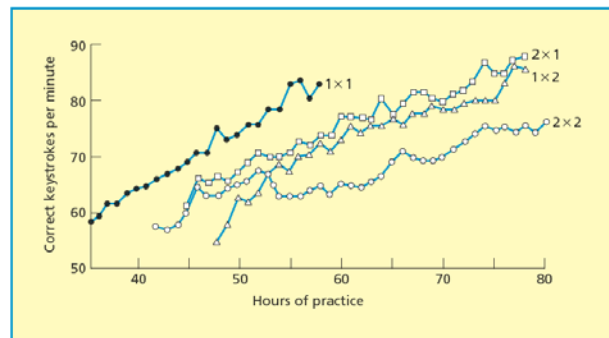
“...with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time.”



Learning curves. Third grade children drilled in addition. The upper curve shows improvement of section working once a day, 10 minutes, for 10 days. The lower curve shows improvement of section working 10 minutes twice a day for 5 days.

(Pyle, 1913)

57

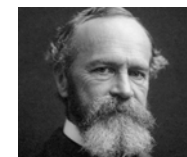


Rate of learning a typing skill for a range of training schedules: 1 × 1 equals one session of 1 hour per day, 2 × 1 equals two such sessions, 1 × 2 is one session of 2 hours and 2 × 2 two 2-hour sessions.

(Baddeley & Longman, 1978)

58

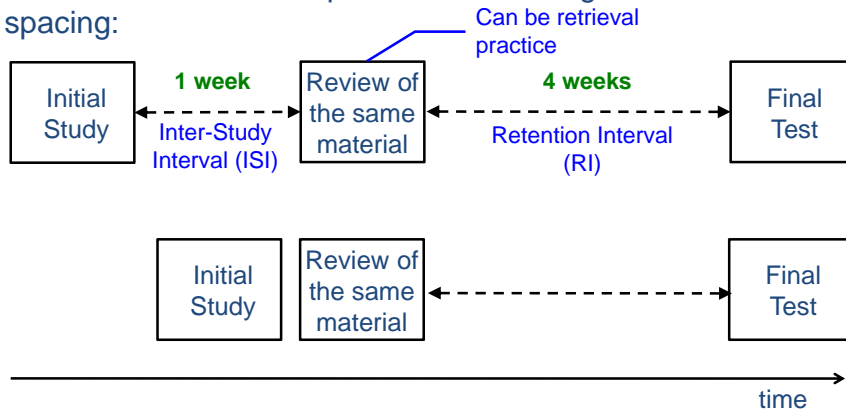
“...You now see why ‘cramming’ must be so poor a mode of study. Cramming seeks to stamp things in by intense application immediately before the ordeal. But a thing thus learned can form but few associations. On the other hand, the same thing recurring on different days, in different contexts, read, recited on, referred to again and again, related to other things and reviewed, gets well wrought into the mental structure. This is the reason why you should enforce on your pupils habits of continuous application.”



- William James, 1899, *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*

The Spacing Effect

Basic structure of an experiment examining the effect of spacing:



Spacing effect: Spaced > Massed

Lag effect: Comparison of different levels of spacing

Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis

Nicholas J. Cepeda
University of California, San Diego and
University of Colorado at Boulder

Harold Pashler, Edward Vul, and John T. Wixted
University of California, San Diego

Doug Rohrer
University of South Florida

Percentage Correct on the Final Recall Test for Massed and Spaced Conditions, Number of Performance Differences and Studies, Total Number of Participants Summing Across All Study/Condition Combinations, and Statistical Analyses, for Spaced Versus Massed Presentations

Retention interval	% Correct			No. of performance differences	No. of studies	No. of participants	Statistical analysis
	Massed	Spaced	SE				
1–59 s	41.2	50.1	1.7	105	96	5,086	$t(208) = 3.7, p < .001$
1 min–less than 10 min	33.8	44.8	1.5	124	117	6,762	$t(246) = 5.0, p < .001$
10 min–less than 1 day	40.6	47.9	8.1	11	10	870	$t(20) = 0.6, p = .535$
1 day	32.9	43.0	6.0	15	15	1,123	$t(28) = 1.2, p = .249$
2–7 days	31.1	45.4	7.3	9	9	435	$t(16) = 1.4, p = .190$
8–30 days	32.8	62.2	8.8	6	6	492	$t(10) = 2.3, p < .05$
31 days or more	17.0	39.0		1	1	43	
All retention intervals	36.7	47.3	1.1	271	254	14,811	$t(540) = 6.6, p < .001$

Cepeda et al. (2006)

August 1988 • American Psychologist

The Spacing Effect

*A Case Study in the Failure to Apply the Results
of Psychological Research*

Frank N. Dempster University of Nevada, Las Vegas

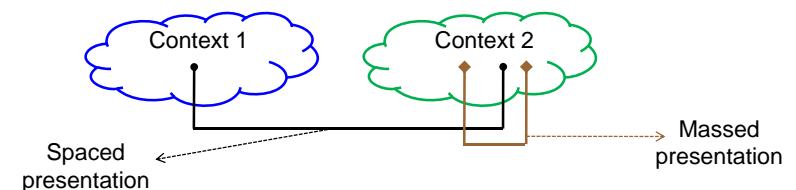
“The spacing effect would seem to have considerable potential for improving classroom learning, yet there is no evidence of widespread application...

Even experienced educators, when judging the instructional effectiveness of text passages, tend to rate prose in which the repetition of a given unit of information is massed as better than those in which it is spaced...

In short, the spacing effect is neither intuitively obvious, nor well known among educators.”

Theoretical accounts •

- Encoding variability theory
 - Item and its context stored at encoding;
 - Context is assumed to undergo random drift;
 - Average distance between any prior context and the current context will increase with passing of time;
 - Likelihood of successful retrieval depends on the distance between context at test and context at encoding;
 - As ISI increases, increased probability that test context will be similar to at least one of the study/encoding contexts
- Deficient processing theory
 - At short ISI, processing of 2nd presentation is deficient; less attention paid to an item that is relatively more familiar



Theoretical accounts

- Study-phase retrieval theory
 - A restudy/review opportunity is more efficacious when the 2nd (or subsequent) encounter with an item *reminds* the learner of the previous encounter
 - i.e., an automatic study-phase retrieval
 - The benefit is greater the more effortful the study-phase retrieval is, which explains the advantage of spacing

Applications of Testing & Spacing in Tutoring Software

- Supermemo

www.supermemo.com

- Spaced Ed

www.spaceded.com

WIRED MAGAZINE: 16.05

Want to Remember Everything You'll Ever Learn?
Surrender to This Algorithm

By Gary Wolf 04.21.08

