The Testing Effect: Basic Research and Educational Application

Neil W. Mulligan

UNC - Chapel Hill

The Testing Effect

- We usually use tests to evaluate
  - That is, we usually think of tests as measurement tools for assessing knowledge or learning
  - But these tools have the unusual property of affecting what they are intended to measure

- The testing effect
  - Direct testing effects
    - Memory retrieval often enhances later memory
      - For the material originally tested
      - For related material that was not tested (transfer)
  - Indirect testing effects
    - A looming test may impact motivation
    - Taking a test can impact new learning (the forward testing effect)
      - New information
      - Restudy of previous information

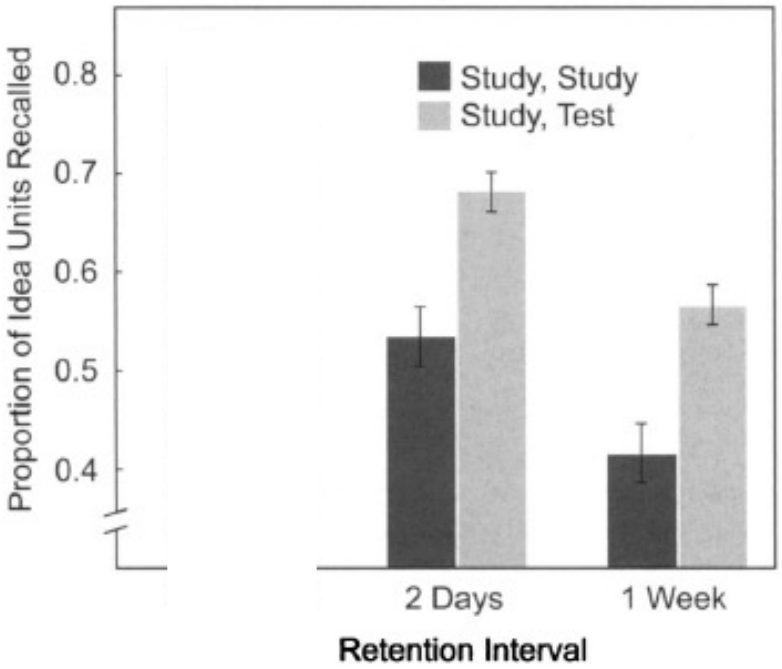
- "Testing" here refers to:
  - Tests
  - Quizzes
  - Activities in which the student retrieves information from memory
  - "Testing" benefits can occur with high stakes, low stakes, and no stakes at all
    - Usual classroom implementation is low or no stakes

Direct Effects of Testing

- Lab studies on the testing effect
  - 1 – Study new materials
  - 2 – Restudy v. Test (retrieval practice)
  - 3 – Final Test

Advantage can be attributed to testing rather than re-exposure

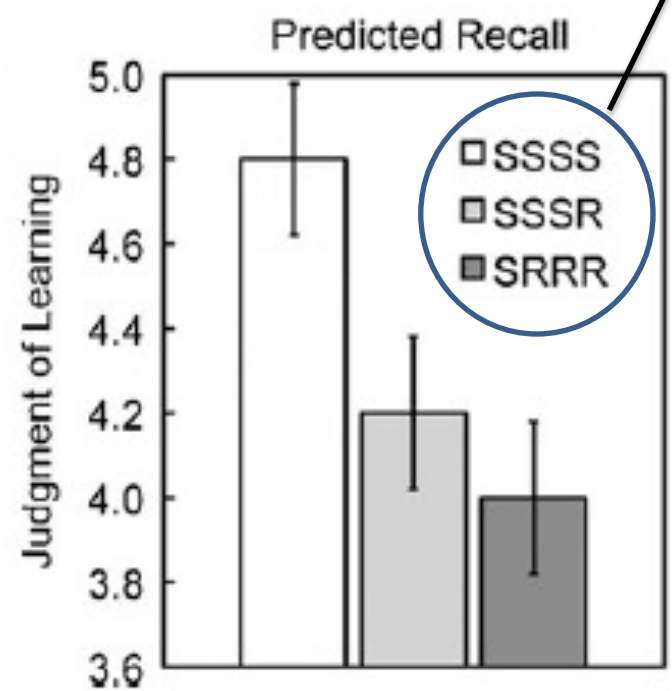
The practice test did not use feedback (and did not produce perfect accuracy), meaning that restudy condition had the advantage of complete re-experience of the study materials



Direct Effects of Testing

Phase 1 = S

Phase 2 =
S S S,
S S T, or
T T T



Summary of Some "Lab" Results

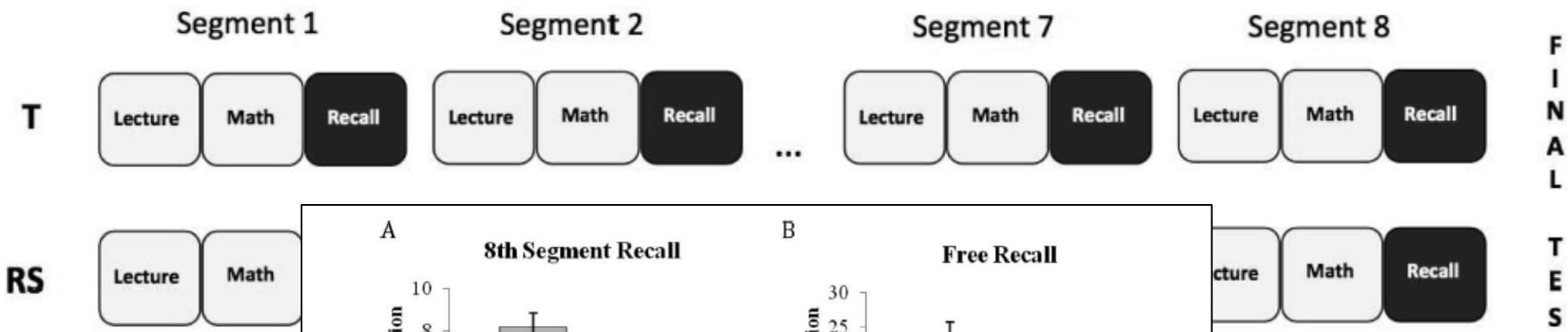
- Direct testing effect for various types of initial (practice) tests and final tests:
  - Free recall
  - Cued recall
  - Recognition
    - Multiple choice

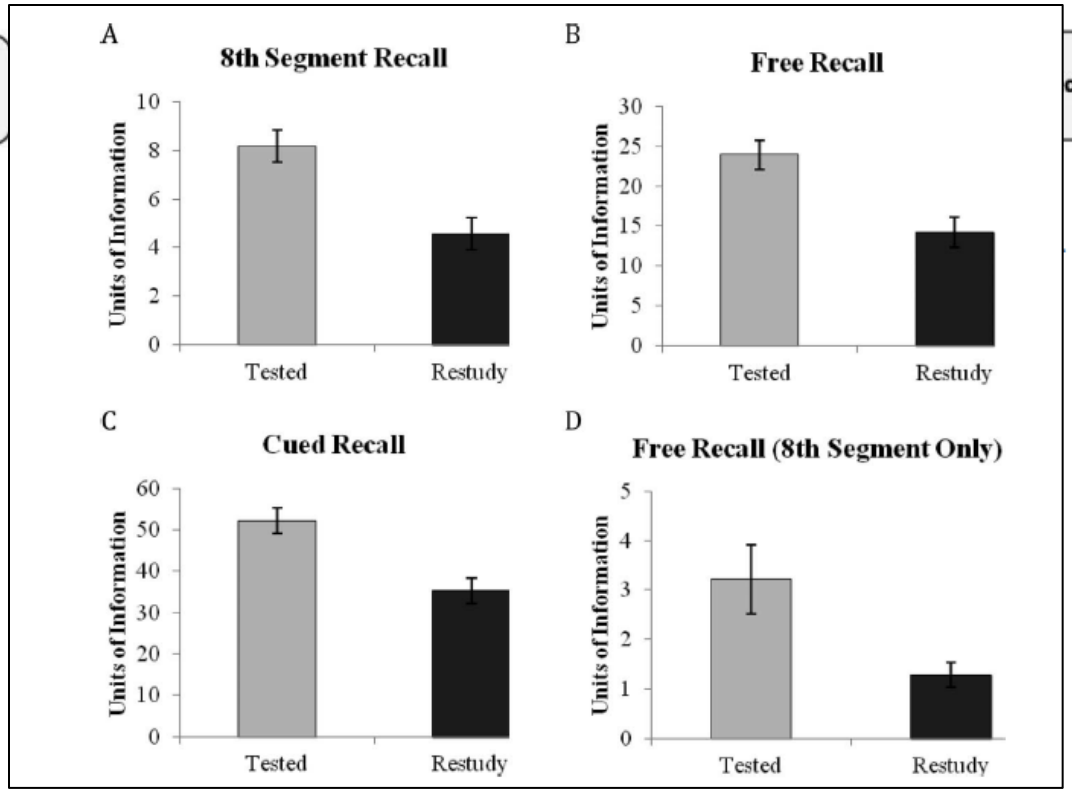
- Populations
  - College students
  - Across the adult lifespan (at least into the late 70s)
  - Memory-impaired populations
  - Children (at least as young as first grade)
  - Online samples
  - Test anxiety

- Role of initial retrieval level
  - Small effect for low initial retrieval (w/out feedback)
  - Robust effect for higher levels of initial retrieval (w/out feedback)
    - Approx. dose response function
  - Testing effect after short retention interval with high levels of retrieval (w/out feedback)

- Feedback
  - Generally, a robust testing effect when retrieval practice includes feedback
    - Regardless of initial retrieval level
    - And after short or long retention interval
  - Special concern about multiple choice practice tests
    - Lures can be incorrectly remembered as correct
    - Importance of corrective feedback

Indirect Effects of Testing: The Forward Testing Effect

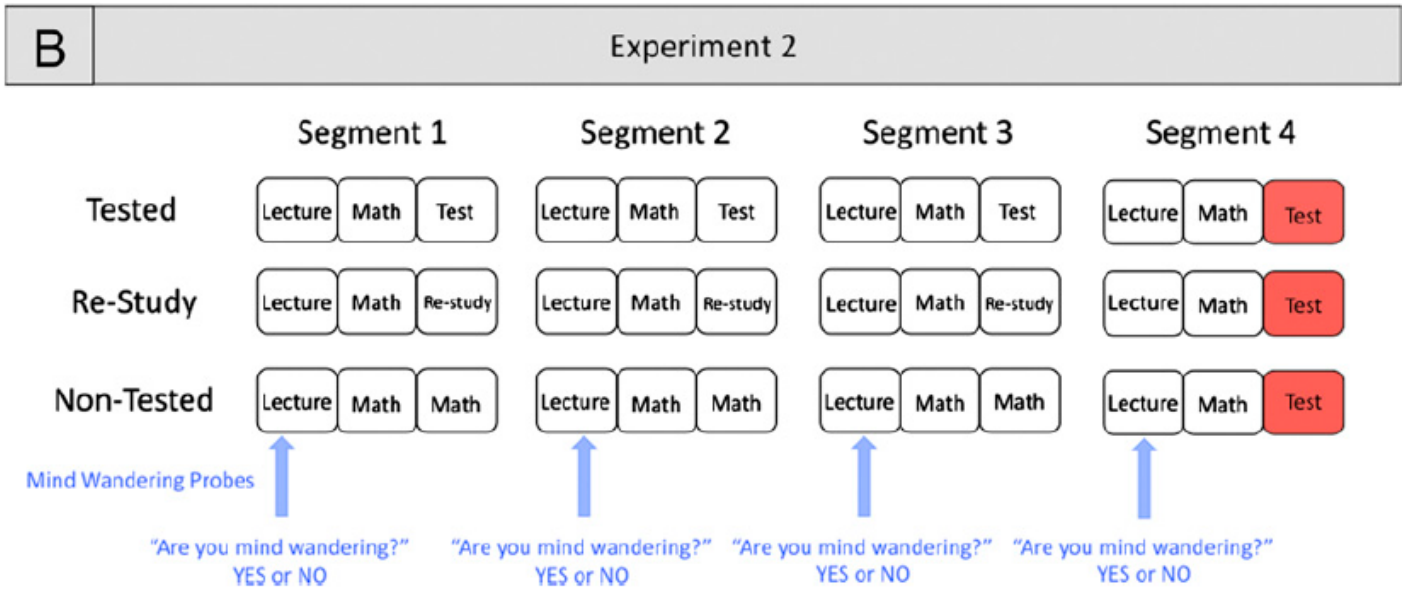


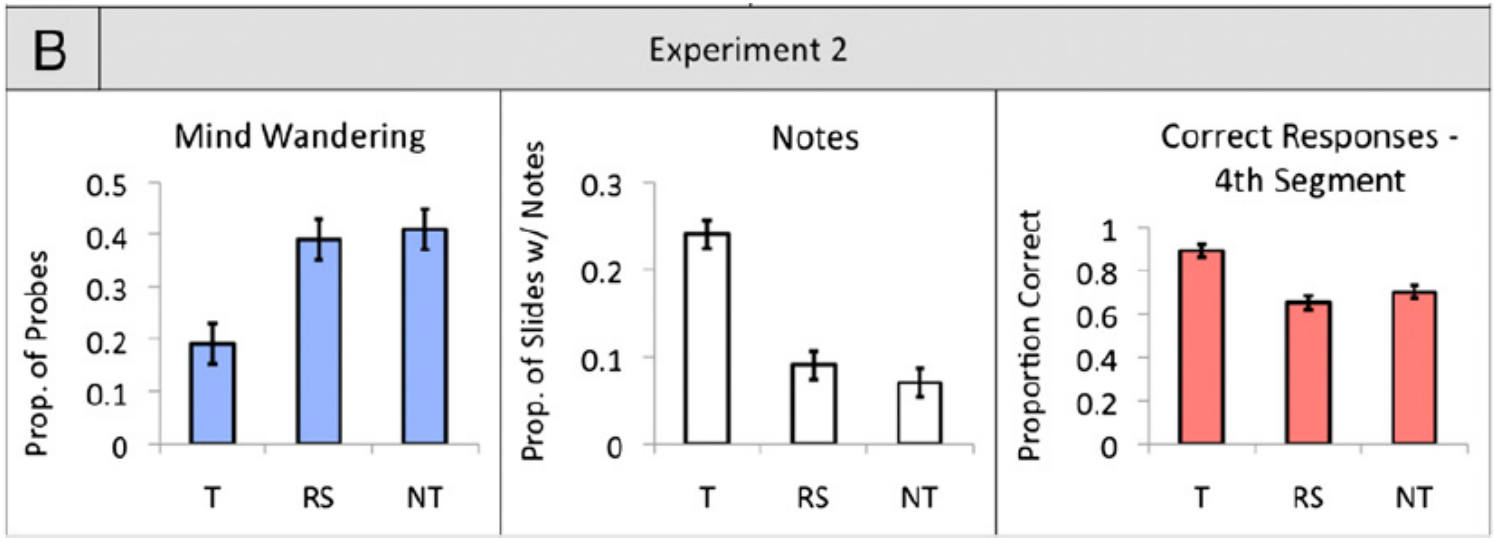


Jing et al (2016)

Indirect Effects of Testing: Mind Wandering and Note Taking

Szpunar et al (2013)





Testing Effects in the Classroom

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

http://dx.doi.org/10.1037/bul0000309

Testing (Quizzing) Boosts Classroom Learning: A Systematic and Meta-Analytic Review

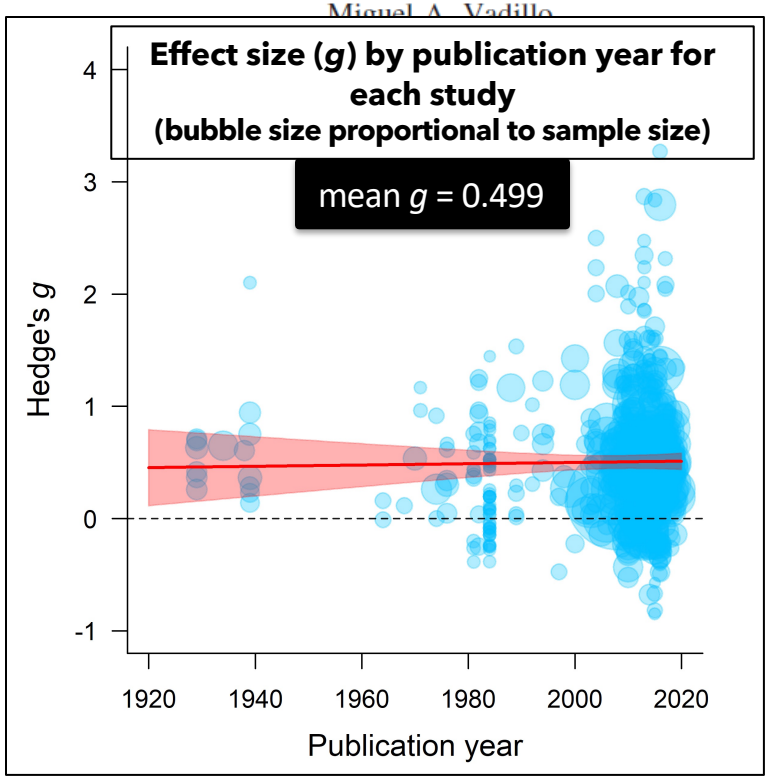
Chunliang Yang and Liang Luo
Beijing Normal University

Rongjun Yu
National University of Singapore

Miguel A. Vadillo

222 published and unpublished studies

573 effects, N = 48,478



Testing Effects in the Classroom

Single-class design

Conditions	Acquisition phase	
Test	Study	Initial test
Control	Study	Control

[short (e.g., 5 min) or long (e.g., 2 weeks) retention interval]

Final test

Multi-class design

Conditions	Class 1		Class 2		Class 3…	Final class	
Test	Study	Quiz	Study	Quiz	...	Study	Quiz
Control	Study	Control	Study	Control	...	Study	Control

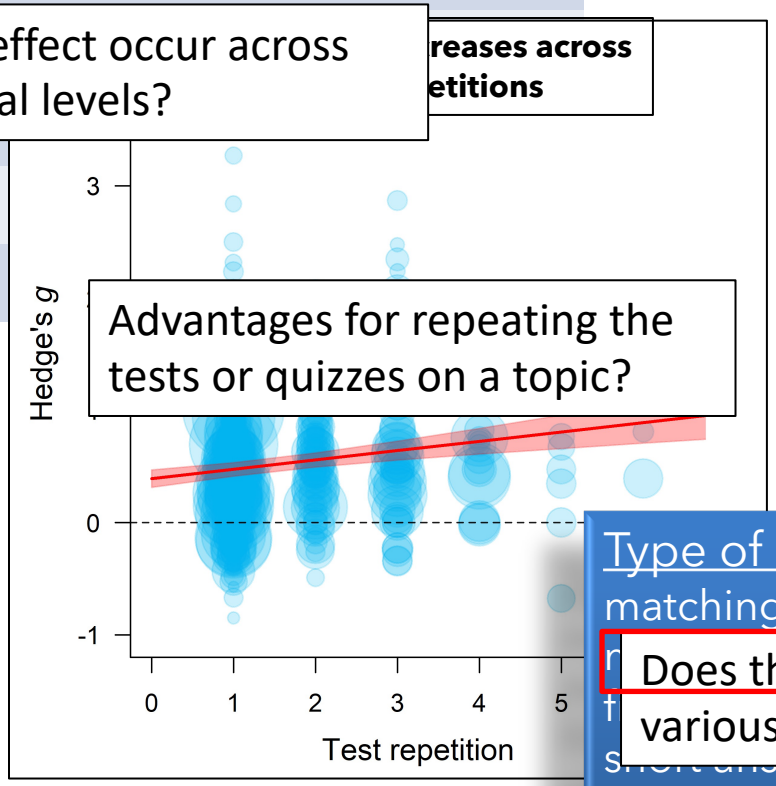
Semester or academic year exam

The Multi-class designs combine direct testing effects and forward testing effects. There are likely multiple components to the observed "testing effect" in these studies.

Testing Effects in the Classroom

	#	*g*	95% CI
Elementary school	43	0.328	[0.085, 0.571]
Middle scho			
High schoo			
University/College			
Continuing education			

Does the effect occur across educational levels?



Advantages for repeating the tests or quizzes on a topic?

Potentially good news! Perhaps we don't have to worry too much about MC quizzes.

Type of Quiz	g
matching	$g = 0.913$
cued recall	$g = 0.316$
free recall	$g = 0.238$

Does the effect occur for various test or quiz types?

Testing Effects in the Classroom

- Control Activity:
  - Robust effect relative to:
    - restudy
    - testing with fewer questions
    - no filler activity
  - But effect is larger for no-filler activity – indicating that some of the apparent testing effect is probably not due to retrieval practice, per se, but simply due to additional processing of the information
  - Elaborative strategies (e.g., concept mapping, note taking, summarizing, etc.)
    - Smaller effect
    - But still a testing advantage
  - Consistent with lab studies

- Corrective Feedback enhances testing effect
  - … but the effect still occurs in the absence of feedback

- Testing effects reported in many disciplines

- Generalizes across knowledge types
  - Fact knowledge, application (e.g., problem-solving), conceptual knowledge

- Some specificity to the testing effect
  - Larger effects when quizzes and final tests have the same format

- Stake level
  - Similar testing effects for high and low stake quizzes (or tests)
  - Multiple low-stake quizzes reduce test anxiety on later exams
    - Not so for high-stake quizzes

Some additional questions/issues

- Student beliefs about tests
  - Survey results on students use and understanding of self-quizzing
    - Far more likely to report highlighting, re-reading, etc.
    - Use self-quizzing to diagnose knowledge state rather than as a learning activity
    - Low use due to metacognitive failure?
      - Perhaps due to delayed benefits of testing (recall the earlier study on memory predictions)
      - "failure" in a limited sense – the belief may be accurate about current accessibility of information
  - Surveys of instructors' beliefs

- Lingering questions about the type of information or knowledge that is affected:
  - Deductive Inference
  - Formation of complex knowledge structures
    - Transitive inference:
      - A>B; B>C; C>D; …
      - Learn via restudy or retrieval practice
      - Final test: B compared to D ?
      - Generally negative effect of retrieval practice (a neg testing effect)
    - Schemas abstraction?
  - Near vs. far transfer

- Does testing benefit all students?
  - Cognitive ability (e.g., working memory capacity)
  - Level of prior knowledge
    - Correlational studies are all over the place
    - Need a good experimental analysis (Zach Buchin)

- Regardless, we shouldn't lose sight that testing clearly works in many ways

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710.

Chan, J. C., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin, 144,* 1111–1146.

Yang et al. (in press). Testing (Quizzing) Boosts Classroom Learning: A Systematic and Meta-Analytic Review. *Psychological Bulletin.*