Turning Tests into Desirable Difficulties: How to Assess Learning in Ways that Enhance Learning

Elizabeth Ligon Bjork
University of California, Los Angeles, CA

Science of Learning Research Symposium
Center for Teaching and Learning
Columbia University
October 11, 2018
Primary Collaborators

Robert Bjork
UCLA

Ben Storm
UC Santa Cruz

Jeri Little
CSU, East Bay

Erin Sparck
UCLA

Nick Soderstrom
LastingLearning.com
Overarching Goals

• Power of tests as tools of learning
• Better understanding of how—in general—the human memory system works
  • In particular, that
    » Performance does not equal learning
    » Immediate assessments tend to measure performance—not learning
    » Forgetting plays an important role in learning
• Familiarity with framework of the New Theory of Disuse (NTD) & concept of “desirable difficulties”
• When and how to use tests to optimize students’ long-term learning and transfer
• We want to use conditions of instruction that optimize learning

• Challenging because the main resources we have to determine what those conditions are
  
  • intuitions/common sense
  
  • observations of what seems to be working in the classroom
    
    • even what our students are telling us

• All turn out to be poor guides for making the correct decision
We—and our students are easily misled into thinking that conditions of instruction/practice that make performance improve rapidly—must be optimal for learning.

Problem: Typically do not support long-term learning and transfer.

Instead: Conditions of instruction that present challenges or difficulties to learner—even appearing to slow rate of learning—often optimize long-term retention and transfer.
Examples of Learners being Misled

**Generation**

- Predicted vs. Actual Percent Recall
- 1st, 2nd, 3rd, 4th Response Time Quartile

**Interleaving Skills**

- Predicted vs. Actual Percent Target Time
- Blocked Practice vs. Random Practice

**Spacing**

- Satisfaction vs. Speed of Learning
- Massed vs. Spaced Spacing of Practice

**Pre-testing**

- Judged vs. Actual Mean Rank
- Read (Long) vs. Pre-test

**Interleaving Concepts**

- Judged vs. Actual Proportion of Participants
- Blocked vs. Interleaved

**Stability Bias**

- Predicted vs. Actual Percentage Correct
- 1 to 4 Trial
Desirable Difficulties

• We call such conditions “desirable difficulties.”

• They are difficulties because they pose challenges for learners and appear to slow the rate of learning.

• They are desirable because they enhance our ability to perform later, when it really matters.
Learning versus Performance

• Bottom line—and what we must understand as teachers is that
  – What we observe is performance
  – But what we must infer is learning
  – And the former (i.e., performance) is an unreliable indicator of the latter

• To combat this problem, need to use desirable difficulties in our teaching
Desirable Difficulties

• Varying the conditions of learning (e.g., Where you study; What you study)
• Distributing or Spacing out repeated study or practice sessions
• Providing “contextual interference” during learning
  – (e.g., Interleave rather than block practice)
• Using tests (rather than re-presentations) as learning events
Desirable Difficulties

• Varying the conditions of learning (e.g., Where you study; What you study)

• Distributing or Spacing out repeated study or practice sessions

• Providing “contextual interference” during learning
  – (e.g., Interleave rather than block practice)

• Using tests (rather than re-presentations) as learning events
Benefits of Testing/Retrieval Practice

• Retrieving information or procedures is a powerful learning event
  – Info/procedures retrieved become more recallable in future than would have been otherwise;
    • Retrieving substantially more powerful than re-presentation as a learning event)

• Tests provide better feedback as to what has or has not been learned/understood (vs. presentations)

• Tests (even failed tests) potentiate the effectiveness of subsequent study opportunities
Benefits of Testing/Retrieval Practice
(Roediger & Karpicke, 2006)

• Participants studied a to-be-learned passage on the sun or on sea otters (about 30 idea units per passage)

• Two learning conditions of interest
  – SSSS: four consecutive 5-min study periods
  – STTTT: one study period plus four consecutive tests of recall for the passage (no feedback)

**Study**

- Read entire passage 14.2 times
- Procedure:
  - 5 mins Study

**Study**

- Read entire passage 3.4 times
- Procedure:
  - 5 mins Study

**Test**

- *No feedback given during tests*
Results

Proportion Recalled

Retention Interval

SSSS

STTT

5 mins

1 week

Retention Interval

Proportion Recalled
Results

Proportion Recalled vs Retention Interval

- **Proportion Recalled**: 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00
- **Retention Interval**: 5 mins, 1 week

Graph showing data for SSSS and STTT conditions.
Proportion Recalled

Retention Interval

Rate: How much will you remember in one week’s time?

SSSS

STTT

SSSS

STTT
But What about Multiple-Choice Testing?

• Do such benefits also arise with M-C testing?
• Widespread opinion: No
  – M-C tests seen as not engaging the kind of retrieval processes needed to support long-term retention
    • The correct answer presented among alternatives, thus bypassing need for retrieval
    • Instead, test-takers led to rely on recognizing correct answer
  – Some studies had shown that tests involving recognition less effective than tests involving retrieval for enhancing long-term retention (e.g., Carpenter & DeLosh, 2006; Glover, 1989)
But is This an Intrinsic Property of Multiple-Choice Testing?

• Or can MC tests be written so as to trigger productive retrieval processes?
  – Not only as to why a correct alternative is correct, but also why an incorrect alternative is incorrect?

• If so, then such MC tests could be particularly good as practice quizzes or tests
  – On criterion tests, identical questions unlikely to be asked
  – Whereas related questions for which previous incorrect alternatives might now be the correct answer are more likely to be asked
General Paradigm for Examining this Question

Study Passage 1: Yellowstone Park

Study Passage 2: Solar system

Cued-Recall Test from one Passage (10 questions)

or

Multiple-Choice (MC) Test from one Passage (10 questions)

Filler Task – 5 min

Final Test: Cued-Recall w/ previously tested, related, & control questions
Practice Question:

The area between the A ring and the B ring is known as the ____.
A. Cassini Division
B. Encke Gap
C. Maxwell Gap
D. Roche Division
Practice Question:

The area between the A ring and the B ring is known as the ____.

A. Cassini Division
B. Encke Gap
C. Maxwell Gap
D. Roche Division
Practice Question:
The area between the A ring and the B ring surrounding Saturn is known as the _____.
A. Cassini Division
B. Encke Gap
C. Maxwell Gap
D. Roche Division

Final Test Related Question:
The area between the A ring and the F ring surrounding Saturn is known as the ______.
Roche Division
Pairs of Related Practice and Final-Test Cued-Recall Questions

• About the same topic (e.g., geysers in Yellowstone Park)
  – With same alternatives (e.g., Old Faithful, Steamboat Geyser, Castle Geyser, Daisy Geyser)
  – But with different correct answers

• Example:
  – Practice: What is the tallest geyser in Yellowstone National Park?  Answer: Steamboat Geyser
  – Final test: What is the oldest geyser in Yellowstone National Park?  Answer: Castle Geyser
Results: Final Cued-Recall Test for Previously Tested Questions

The graph shows the correct performance percentages for previously tested questions, broken down by initial test type. The bars indicate that the cued-recall test type had a higher correct performance percentage compared to the multiple-choice test type. The error bars suggest variability in performance.
Results: Final Cued-Recall Test for Related (not previously tested) Questions

Correct Performance Percentages

- Related
- Control

Initial Test Type
- Multiple-Choice
- Cued-Recall
General Research Strategy in Classroom

- Give MC Quizzes over duration of large undergraduate research methods course (10-wks)
- On Final Exam: Three Types of Critical Questions (5 each)
  - IDENTICAL REPEAT (Had appeared on a previous quiz)
  - CONCEPTUAL REPEAT (Related to Question appearing on a previous quiz, but not itself previously tested)
  - CONTROL ITEMS
- IR & CR Items Counterbalanced Across Two Terms

<table>
<thead>
<tr>
<th>TERM 1</th>
<th>TERM 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDENTICAL REPEAT</td>
<td>CONCEPTUAL REPEAT</td>
</tr>
<tr>
<td>CONCEPTUAL REPEAT</td>
<td>IDENTICAL REPEAT</td>
</tr>
</tbody>
</table>
General Pattern of Results from Classroom (Bjork, Little, & Storm, 2014, *JARMAC*)

- For IDENTICAL REPEATS
  - Performance significantly improved from Quiz (82%) to Final Exam (92%)
  - Performance significantly > than that for CONTROL items
- For CONCEPTUAL REPEATS
  - Performance (90%) significantly > than that for CONTROL items
General Conclusions

• MC testing format can be a win-win condition for practice quizzes
  – Can foster test-induced learning not only of previously tested information, but also of related information
  – Particularly important feature, given that items on practice quizzes rarely repeated verbatim on later criterion exams.
Why the benefit for related information following an initial MC test?

Hypothesis:

- Participants attend to all alternatives in order to eliminate incorrect choices.
- In doing so, they may retrieve—and thereby strengthen—information from the studied passage that pertains to an incorrect alternative.
- Should this information then become the basis for a related question on a later test, one’s ability to access answer should be improved.
Testing the Hypothesis
(Little & Bjork, 2015)

• If so, the competitiveness of the incorrect alternatives should matter

• Following study of passages, participants received MC tests containing both competitive & noncompetitive questions

Example Questions with Both Competitive and Non-Competitive Alternatives

<table>
<thead>
<tr>
<th>Question</th>
<th>Correct Answer</th>
<th>Competitive Alternatives</th>
<th>Noncompetitive Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Which outer planet was discovered by mathematical predictions rather than by direct observation?</td>
<td>Neptune</td>
<td>Saturn, Uranus</td>
<td>Mercury, Mars</td>
</tr>
<tr>
<td>(b) What is the hottest terrestrial planet?</td>
<td>Venus</td>
<td>Mercury, Mars</td>
<td>Saturn, Uranus</td>
</tr>
</tbody>
</table>
Results

Correct Recall Percentages

- Competitive Alternatives
- Non-Competitive Alternatives
- Control

Related
Summary of Results So Far
(Little et al., 2012; Little & Bjork, 2015; Bjork, Little, & Storm, 2014)

On Cued-Recall Final Tests,

- MC practice quizzes enhanced performance for repeated questions equally as well (and sometimes better) than did cued-recall practice quizzes.
- Moreover, MC practice quizzes enhanced performance for related questions compared to matched control questions; whereas cued-recall quizzes did not (and sometimes impaired performance).
- Pattern for related questions remained same with or without feedback on MC practice quiz.
- Benefits of MC quizzes for related questions last over (at least) 48 hours.
- In classroom, have found benefits for related questions to last up to 10 weeks—that is, to end of course.
- In sum, the MC Format—when used properly—can be a Win-Win condition for practice quizzes.
Next Research Question: Can MC Pretests Function as a Desirable Difficulty?

• Know from previous research that CR pretests can improve later recall for tested information? (e.g., Richland, Kornell, & Kao, 2009)

• Would this also be true for MC pretests?

• And, what about effect on related information?
  • A CR pretest might direct attention to pretested information, but reduce attention directed towards competitive non-pretested information
  • In contrast, a MC pretest might direct attention more broadly

• Have looked at this question in both the laboratory and the classroom.
Before reading passage (e.g., Solar System, Yellowstone, Stimulants),

• Ss given either an MC or a CR pretest

• Then, compared their performance on a Final Cued-Recall test against when they engaged in other types of pre-study activities, such as
  ✦ Studying facts
  ✦ Extra-time to study passages
How many milligrams of caffeine are contained in a Jolt Cola?

A. 96
B. 280
C. 300
D. 414

Before reading passage about stimulants,

“Jolt Cola has a lot of caffeine”

“96 seems low”

“414 seems too specific”

“Maybe this one is right”
...as of today, Netherlands is the leading nation with regard to consumption of caffeine per person. The average person consumes 414 mg in one day...Today in the U.S., about 80% of adults consume caffeine daily, although their daily intake is closer to 168 mg... The amount of caffeine in any beverage varies: a Coca Cola contains 96 mg, Jolt Cola contains about 280 mg, and a cup of brewed coffee contains about 100 mg. of caffeine... An acute overdose of caffeine, usually in excess of about 300 mg, can result in a state of CNS over-stimulation called caffeine intoxication.
General Pattern of Results: Final CR Test for Related Information

Correct Performance Percentages

- Related to Pretested
- Related to Studied Facts
- Extended-Study

Pre-reading Activity Type

Multiple-Choice/Facts with Competitors
Cued-Recall/Facts without Competitors

Little & Bjork, 2016
Pretesting in the Classroom

– Might pretesting students on material before a lecture facilitate the learning of that material when it is subsequently discussed in the lecture?

– What about for related, but non-pretested, information?
Pretesting in the Classroom

• Pretesting:
  – 300+ students from a psychology research methods course
  – 4 item, multiple-choice pretests were administered immediately before some lectures, but not others
  – Pretested topics were discussed in the following lecture
  – As well as other topics

• Cumulative Final:
  – Three types of questions:
    1. **Identical** to pretested
    2. **Related** to pretested topics, but not itself pretested
    3. **Control** (not pretested and unrelated to pretested topics)
Pretesting in the Classroom

Results:

On Final Exam

– Both Identical and Related questions answered correctly significantly more often than were corresponding control questions.

Soderstrom & Bjork, ms in preparation
Next Questions of Interest

Thinking we now knew the underlying strategy that allows MC tests to enhance performance on later related questions,

- Wondered about extent to which students spontaneously engage in this behavior
- Basis of interviews, seems only ~ 30% of students report ever using strategy

Given these findings, wanted to see how we might increase this percentage

- Explicit Instructions
- New Type of Multiple-Choice Format
Explicit-Instruction Study

Design:
- Half Ss given explicit instructions in how to use “productive” strategy in addition to standard instructions
- Other half: Just standard instructions

Explicit Instructions used:
“When you answer each multiple-choice question, please try to think not only about why your chosen answer is correct, but also why the other alternatives are incorrect. Please follow this instruction even if you know the correct answer without thinking about the alternatives.”
Explicit Instruction Study: Results

Strategy-Instructed Group

• Showed greater improvement on related questions (vs. control questions) than Ss in uninstructed group.
• Some Ss reported not following instructions for all questions (requires more cognitive effort).
• Ss—reporting using strategy most of the time—performed better than those who reported not using it most of the time.

Uninstructed Group

• Post-experiment, Ss asked if ever used such a strategy—only ~1/3 reported ever using it spontaneously.
• Interestingly, some Ss remarked that it sounded like “a good thing to do.”
Alos Explored Effects of a Different MC Testing Format

• MC format developed by James Bruno (1986) called Confidence Weighted Multiple-Choice

• Format in which test takers are encouraged to select answer by assessing their confidence in that answer relative to other alternatives

• Our hypothesis was that this type of format might thus encourage test takers to engage in productive retrieval processes without need for explicit instructions
What is most popular ice cream flavor?
Example Question: Solar System Passage

*What planet lacks an internal magnetic field?*

Venus

Don't know

Mercury

Jupiter
What is the capital of British Columbia, Canada?
Experiment 1

- Using the Little et al. (2012) materials (i.e., Solar System & Yellowstone Passage)

- Compared the benefit of MC testing for related information on a Final Cued-Recall Test using
  - A standard MC test versus
  - A confidence-weighted MC test
Three Conditions: Standard M-C test; Confidence-Weighted M-C test; or Study Only

• Passage 1: Yellowstone or Saturn (9 min)
  Initial Test or Tetris (25 s/question):

• Passage 2: Saturn or Yellowstone (9 min)
  Initial Test or Tetris (25 s/question):

• Tetris Distractor (5 min)

• Final Cued-Recall Test (related information ONLY) Self-paced
Example of an Initial-Test Item

What is the tallest geyser in Yellowstone National Park?

Final test: Related item (cued recall)

What geyser is thought to be the oldest in the world?
Experiment 1: Results

Correct Responses (%)

- No Test
- Standard Multiple-Choice
- Confidence Weighted Multiple-Choice

Post Study Activity

Sparck, Bjork, & Bjork (2016)
Experiment 1: Results

• So, Ability to answer Related Questions on the Final Cued-Recall Test was greater following a CWMC Practice Tests than a Standard Multiple-Choice Practice Test.

Is this due to something special about the CWMC Format (as we suspect)?

What if confidence judgments were taken as part of the Standard MC format?

Would we then see the same benefit for Standard MC format?
Experiment 2

• To test this possibility, conducted a second experiment using the same materials

• But now, compared the benefit for related information on a Final Cued-Recall Test after the following three types of practice tests
  ✩ Standard Multiple-Choice
  ✩ Standard Multiple-Choice + Confidence Judgment (After selecting answer, S asked “On a scale of 0-100 (where is 0 is not at all and 100 is completely), how confident are you in your answer?"
  ✩ Confidence-Weighted Multiple-Choice
Correct Responses (%)

Post Study Activity

- Standard Multiple-Choice
- Standard Multiple-Choice Plus Confidence Judgment
- Confidence-Weighted Multiple-Choice
Wondered if experiencing a confidence-weighted practice test might lead students to transfer a similar test-taking strategy to a standard multiple-choice practice test?

Do students become more sensitive to benefits of retrieving why incorrect alternatives are incorrect after experiencing confidence-weighted testing?

Might that then lead to their using such a strategy on subsequent standard-MC tests?
Experiment 3: Design

• Two Groups

• Following reading of the First Passage (Solar System or Yellowstone), received either:
  – Standard MC Practice Test or
  – CWMC Test

• Then, after reading second Passage, both groups given a Standard MC Practice Test

• Then, after filled retention interval, both groups given a Final Cued-Recall Test for Related Items
Experiment 3: Procedure

Passage 1 (Saturn)
- What planet lacks an internal magnetic field?
  - a) Venus
  - b) Mercury
  - c) Jupiter

Passage 2 (Yellowstone)
- What is the tallest geyser in Yellowstone National Park?
  - a) Castle Geyser
  - b) Steamboat Geyser
  - c) Old Faithful

Tetris
- What is thought to be the oldest geyser in Yellowstone National Park?
  - [Blank]
Proportion Correct on Test 2

Initial Test Type: Passage 1

- Standard Multiple-choice
- Confidence-weighted Multiple-choice
Some Metacognitive Findings

• Generally, participants report liking to take CWMC tests, citing—in particular—that they liked that such tests allowed them to demonstrate partial knowledge

• Most participants report recalling info about incorrect alternatives when answering CWMC questions, but also report that they don’t think doing so would be a helpful learning strategy

• Consistent to results from many other studies, indicating that our students (and often ourselves as well) are not sensitive to what are and are not effective conditions of learning
When properly constructed (i.e., with competitive choices), MC tests can function as an effective tool for learning or as a *desirable difficulty*

- Have an advantage over cued-recall tests in that
  - Not only is retention of tested information benefited
  - So is the retention of related information
  - Whether used as a pretest or a posttest

- Apparently, CWMC tests (also with competitive alternatives) seem to function even more effectively

- Experience with CWMC tests, however, can then boost effectiveness of taking Standard MC tests

- Preliminary results indicating that a Matching Test Format (when properly constructed) may also have similar benefits
Advice for Instructors & Students

Instructors:

• In constructing MC tests, need to use competitive incorrect alternatives to invoke beneficial retrieval processes in our students.

• But may need to train students in how to use strategy.

• Via instructions

• Or by giving them some experience with the CWMC testing format
Concluding Advice for Instructors & Students

Students:

• In taking MC tests—particularly practice quizzes—consider all alternatives, thinking not only why your selected answer is correct but also why other choices are incorrect.
• Do this even when you immediately know (or think you know) correct answer.
• May require extra effort, but worth it—as by so doing, can enhance your later performance—not only on previously tested information—but also on related information.
• Don’t be afraid to take MC tests before studying: Can make your subsequent study more effective—even if initial performance not above chance.
At this point, hope I have convinced you of the value of using tests (or retrieval practice) as a desirable difficulty in your instruction.

Now, want to give you a theoretical framework for understanding the importance of testing and also a feeling for when and how to use tests to enhance long-term memory.

Have called this framework: the New Theory of Disuse (NTD).
Starting observations for NTD
(Bjork & Bjork, 1992)

• Information in memory, no matter how over-learned, becomes inaccessible with long enough period of disuse.
  – (e.g., Street addresses, names of high-school friends, …)

• But that information remains in memory as measured by recognition, priming, and—especially—relearning.

• Called our framework “New Theory of Disuse”

• Thorndike was not all wrong, however:
  – Use (& disuse) do matter—& shape our memories
  – To recognize important role of use in the system, called our framework the New Theory of Disuse
“Important peculiarities” of the human memory system (Bjork & Bjork, 1992)

• A remarkable capacity for storing information coupled with a highly fallible retrieval process.
• What is accessible is heavily dependent on the current cues available.
• Retrieving information from memory is a dynamic process—altering the subsequent state of the system.
• Conditions that produce forgetting—rather than undoing learning—create opportunities for additional learning.
New Theory of Disuse (Bjork & Bjork, 1992)

Starting Assumption:

• Items in memory are indexed by two different strengths: Storage (SS) and Retrieval Strength (RS)
  – SS reflects how well learned an item is; permanent
  – RS reflects accessibility of item given current situational cues
• Importantly, this distinction between SS and RS corresponds, empirically, to learning versus performance
This distinction between SS and RS is by no means new with the NTD; many early learning theorists made such a distinction (to account for various findings).

What is new within the NTD framework is our assumptions as to how SS and RS interact.
How SS and RS are Assumed to Interact in NTD

- Lower the RS, greater the gain in SS from additional study or successful retrieval
- SS acts to retard rate of decrease in RS with additional study or successful retrieval
- This interaction allows the prediction that forgetting (i.e., loss in RS) will enable greater learning (i.e., a greater increase in SS when a study trial or a test occurs after a delay)
Key Implication of this Interaction in NTD

- Conditions that maximally increase RS differ from those that maximize gain in SS.

- Thus, if we (and/or our students) interpret current RS as SS, we can be misled into preferring poorer conditions of instruction to better conditions of instruction.
RETRIEVAL STRENGTH

STORAGE STRENGTH

“Learning”

Low

High

Low

High

? ?

? ?

“Performance”

“Learning” vs. “Performance”

STORAGE STRENGTH

Low

High

RETRIEVAL STRENGTH

Low

High

Hotel room number this weekend


![Diagram showing the relationship between retrieval strength and storage strength.](image-url)
RETRIEVAL STRENGTH

STORAGE STRENGTH

“Learning”

Low

High

Low

High

RETRIEVAL STRENGTH

“Performance”


Childhood phone number

Current phone number

Hotel room number from last year

Hotel room number this weekend

Current phone number

Hotel room number from last year

Childhood phone number
RETRIEVAL STRENGTH

STORAGE STRENGTH

“Learning”

High

Low

Low

High

“Performance”

Childhood phone number

Current phone number

Hotel room number from last year

Hotel room number this weekend

RETRIEVAL STRENGTH

STORAGE STRENGTH

“Learning”

High

Low

“Performance”

Childhood phone number

Current phone number

Hotel room number from last year

Hotel room number this weekend

Implications of the NTD for Effective Testing

• As teachers, we need to delay when we give tests for them to be effective learning events for our students.

• When we give tests immediately or shortly after presentation of the material, what we are measuring is RS or temporary performance—not SS or learning.

• By delaying tests, some forgetting or decrease in RS will occur, allowing such tests to increase SS and, thus, long-term learning.
DEMONSTRATION
1st Recall

5
Grandmother
Vol. Withstanding
Plato
Jefferson
Bridge
Red
Madrid
North
Oak
Rubens

2nd Recall (Cued)

4
Madrid
Theater
Spoon
Red
North
Whale
Oak
Cannon
Tiger
Rubens
Engine
Jefferson
Violet
Grandmother

Book

Plato

Not Withstanding
So, as just demonstrated, another way to decrease RS in order to make testing more effective is to vary the cues (or context) used to test the same material.

Introducing such variation into our test questions will decrease RS at the time of testing, making such questions or tests be more effective learning events (i.e., increase SS of the tested material).

Furthermore, because retrieval is cue (or context) dependent, decreasing RS in this way should increase students’ ability to recall the to-be-learned material given different test cues—which will probably be present on later final exams or end-of-term criterion tests.
Concluding Points

- As teachers, we need to use caution in interpreting current performance, which reflects current RS—not SS or learning.

- If current performance is assumed to index learning
  - Become prone to over-estimating the degree to which learning has been achieved
  - Are led to use poorer testing conditions—such as ones that result in high performance but are based only on RS—and thus not producing long-term learning or SS

- To construct effective tests, useful to be guided by the Goldilocks Principle

- Namely, to test when RS is low—but not too low, so as to make successful retrieval impossible—but not too high, so as to make retrieval too easy and simply a measure of temporary performance.
We have discussed two ways to help you improve the effectiveness of your testing:

1. Delaying your tests so that $RS$ will be lower and, thus, your tests can result in a larger increase to $SS$ or long-term learning.

2. Varying the cues you use in to test the same material,
   - Both will serve to lower $RS$ at the time of the test allowing for an increase in $SS$.
   - Also increase students’ ability to recall the to-be-learned material to a variety of cues—thus, increasing both long-term learning and transfer.

Concluding Points
The End
And, thanks for your attention!
References


