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

222 published and unpublished studies

573 effects, $N = 48,478$
Testing Effects in the Classroom

**Single-class design**

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Acquisition phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Study Initial test</td>
</tr>
<tr>
<td>Control</td>
<td>Study Control</td>
</tr>
</tbody>
</table>

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

**Multi-class design**

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3...</th>
<th>Final class</th>
<th>Semester or academic year exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Study Quiz</td>
<td>Study Quiz</td>
<td>...</td>
<td>Study Quiz</td>
<td>Semester or academic year exam</td>
</tr>
<tr>
<td>Control</td>
<td>Study Control</td>
<td>Study Control</td>
<td>...</td>
<td>Study Control</td>
<td>Semester or academic year exam</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Type of Quiz</th>
<th>Effect Size (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching</td>
<td>g = 0.913</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>g = 0.567</td>
</tr>
<tr>
<td>Fill-in the blank</td>
<td>g = 0.773</td>
</tr>
<tr>
<td>Short Answer</td>
<td>g = 0.638</td>
</tr>
<tr>
<td>Cued Recall</td>
<td>g = 0.316</td>
</tr>
<tr>
<td>Free Recall</td>
<td>g = 0.238</td>
</tr>
</tbody>
</table>

Does the effect occur across educational levels?

Effect size increases 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.

Does the effect occur for various test or quiz types?

Test repetition

Hedge's g

Elementary school

Middle school

High school

University/College

Continuing education
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

