Modeling the Influence of Format and Depth during Effortful Retrieval Practice

Jackie Maass and Dr. Philip Pavlik, Jr.

University of Memphis

Institute for Intelligent Systems

Memphis, TN

What's to Come

- Background on the Testing Effect
- The Current Work
  - Research Questions
  - Experimental Design
  - Results
    - Modeling Retrieval Practice Performance
  - Implications

Introducing…

- Testing Effect
  - Consistent benefit of quizzing oneself as a method of practice over re-reading or re-studying information (e.g., Roediger III & Karpicke, 2006; Thompson, Wegner, & Bartling, 1978)

- Practical Issues for Implementation in Education
  - Ease of administration, item difficulty
    - Question Depth and Answer Format

INSTITUTE for INTELLIGENT SYSTEMS

THE UNIVERSITY OF MEMPHIS

Answer Format

- Benefit of practice tests with less retrieval cues

- Most beneficial formats:

  Free Recall > Cued Recall > Recognition

or Essay Questions > Short Answer > Multiple Choice

(e.g., Glover, 1989; Kang, McDermott, & Roediger III, 2007; McDaniel, et. al., 2007)

Why are less cues better?

- Retrieval effort hypothesis (Pyc & Rawson, 2009)
  - There is more memorial benefit from successful retrieval practice when it is difficult than when it is less difficult
  - Motivated by:
    - Bjork's (1994, 1999) desirable difficulty framework
    - Craik and Lockhart's (1972) depth of processing research

Research Questions

Further investigating the effect of effortful retrieval practice

- Can we increase the difficulty of items through either format (multiple choice vs. short answer) and/or through depth (factual vs. applied)?

- Does retrieval need to be successful in order to be beneficial?

- Do participants benefit from more difficult items?
  - If so, is the cause of the increase in difficulty an important factor (i.e., depth or format)?

My Design

2 (retrieval format: MC, SA)

x 2 (retrieval depth: factual, applied)

Between-subjects
Fully-factorial

x 2 (posttest format: MC, SA)

x 2 (posttest depth: factual, applied)

Within-subjects

MC= Multiple Choice
SA= Short Answer

INSTITUTE for
INTELLIGENT
SYSTEMS

THE UNIVERSITY OF
MEMPHIS

Materials

Text

The heart is a pump. Its walls are made of thick muscle. They can squeeze (contract) to send blood rushing out. [...] The oxygen-rich blood is returned to the left atrium of the heart and pumped out to the body through the left ventricle.
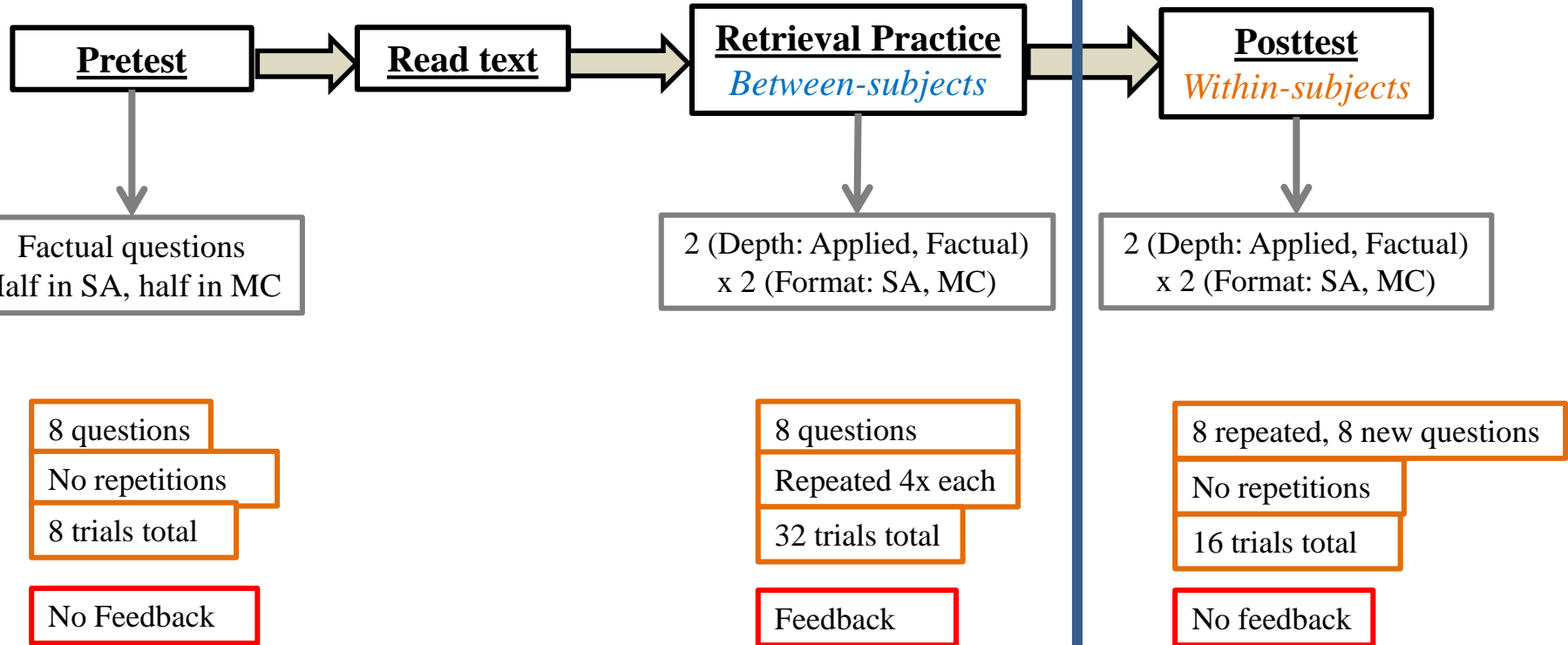
Factual

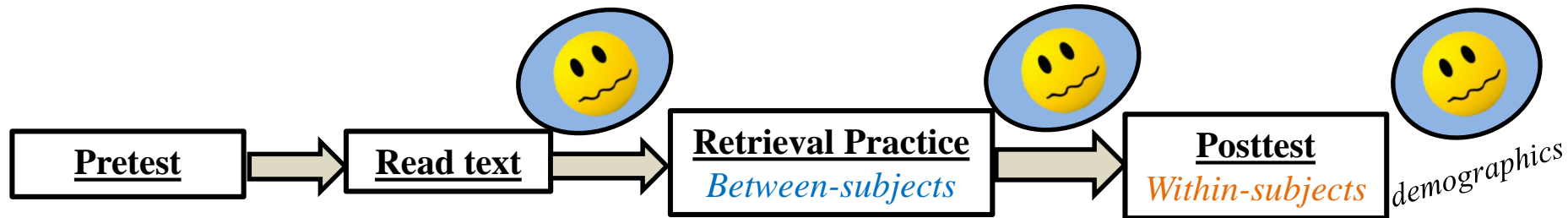Which component of the circulatory system acts as a pump?

Applied

Why doesn't oxygen rich blood flow directly from the lungs to the rest of the body?

INSTITUTE for INTELLIGENT SYSTEMS

THE UNIVERSITY OF MEMPHIS

Procedure

Pretest → **Read text** → **Retrieval Practice** *Between-subjects* → **Posttest** *Within-subjects*

Factual questions
Half in SA, half in MC

2 (Depth: Applied, Factual)
x 2 (Format: SA, MC)

2 (Depth: Applied, Factual)
x 2 (Format: SA, MC)

8 questions

No repetitions

8 trials total

No Feedback

8 questions

Repeated 4x each

32 trials total

Feedback

8 repeated, 8 new questions

No repetitions

16 trials total

No feedback

INSTITUTE *for* INTELLIGENT SYSTEMS

Department of Psychology

THE UNIVERSITY OF
MEMPHIS

Procedure

Pretest → Read text → **Retrieval Practice** *Between-subjects* → **Posttest** *Within-subjects* → *demographics*

Currently I am feeling **frustrated.**

Strongly Agree

Agree

Neutral

Disagree

Strongly Disagree

6 emotions: frustration, anxiety, confusion, discouraged, bored, distracted

Results

- Participants
  - *N* = 178 after 15 were removed
    - (5 glitches, 10 had 10+ timeouts)

  - Amazon Mechanical Turk; $5; 1 hour
  - Native English Speaker/ US or Canada
  - "Reliable" Mturk worker

Modeling Retrieval Practice

- Logistic mixed-effects regression to model performance during the retrieval practice phase

- Based on a Performance Factors Analysis (Pavlik, Cen, & Koediger, 2009)
  - differentiates prior incorrect and correct trials.

Participant Performance during Practice
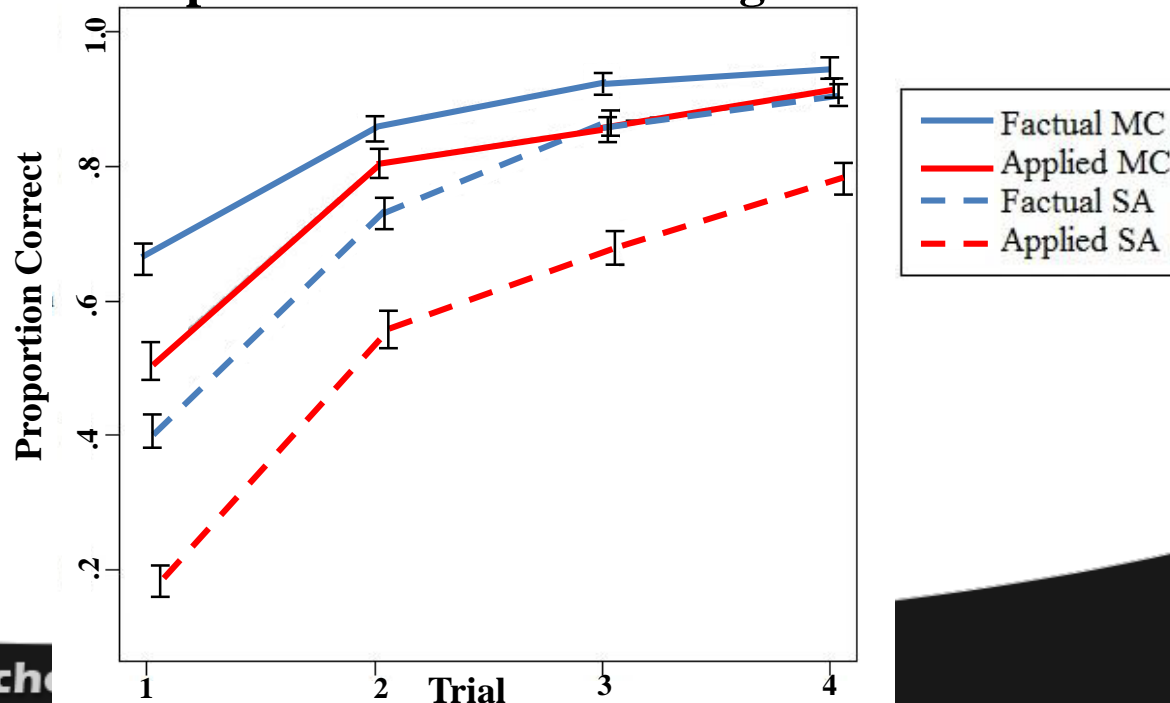
Process of Modeling

- 2 parameters to capture the count of prior correct and incorrect trials

- Pretest score and a random effect of participant

- Variables to capture format, depth, time on task, affect, and interactions of these factors with the (in)correctness of previous trials

- Final Model= 9 parameters, $R^2 = .359$

INSTITUTE for INTELLIGENT SYSTEMS

THE UNIVERSITY OF MEMPHIS

Summary of Fixed Effects for Logistic Regression Model Predicting Future Success

Parameter

Intercept
Pretest
Count of Prior Correct
Count of Prior Incorrect
Format
Depth
Prior Correct x Format
Prior Correct x Depth
Prior Incorrect x Confusion

log of (1 + the prior count)

MC = 0   SA = 1
Factual = 0   Applied = 1

Average of confusion before & after practice

Summary of Fixed Effects for Logistic Regression Model Predicting Future Success

Parameter			
Intercept	-0.11	.19	-.56
Pretest	1.95	.30	6.50
Count of Prior Correct	1.82	.16	11.72
Count of Prior Incorrect	1.47	.15	9.88
Format	-1.22	.14	-8.93
Depth	-0.82	.13	-6.06
Prior Correct x Format	1.13	.19	5.93
Prior Correct x Depth	0.36[†]	.19	1.93
Prior Incorrect x Confusion	-0.18	.05	-3.78

More gained from a successful over unsuccessful trials (Count of Prior Correct / Count of Prior Incorrect)

Negative parameter value for the more difficult format and/or depth (Format / Depth)

More gained from successes on trials of the more difficult format and/or depth (Prior Correct x Format / Prior Correct x Depth)

Note: [†] $p < .05$; all other parameters are significant at the $p < .001$ level.

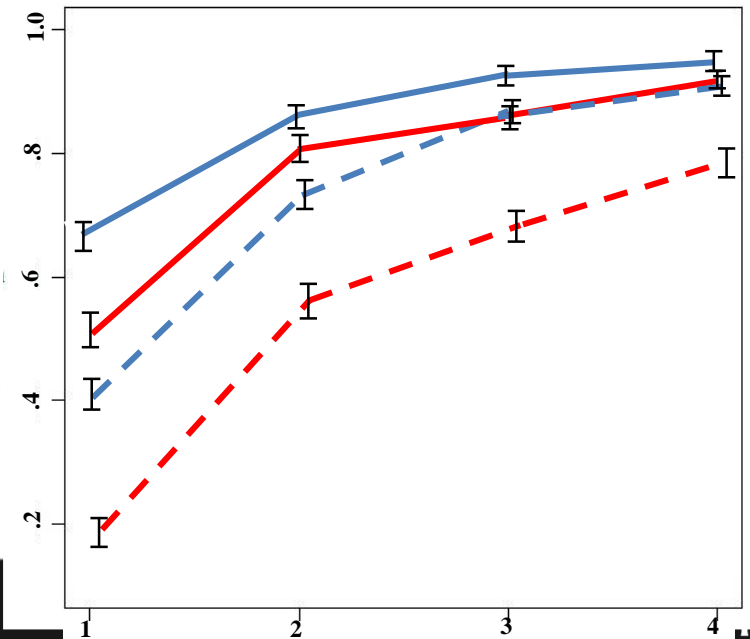*Coding: MC=0, SA=1; Factual=0; Applied=1*

Cross-Validation

- Ten runs of a 10-fold cross-validation
  - Training $R^2 = .293$
  - Testing $R^2 = .284$
- 96.8% of the validity of the model was retained in the held out data

Predicted Performance

Actual Performance

Legend:
- Factual MC
- Applied MC
- Factual SA
- Applied SA
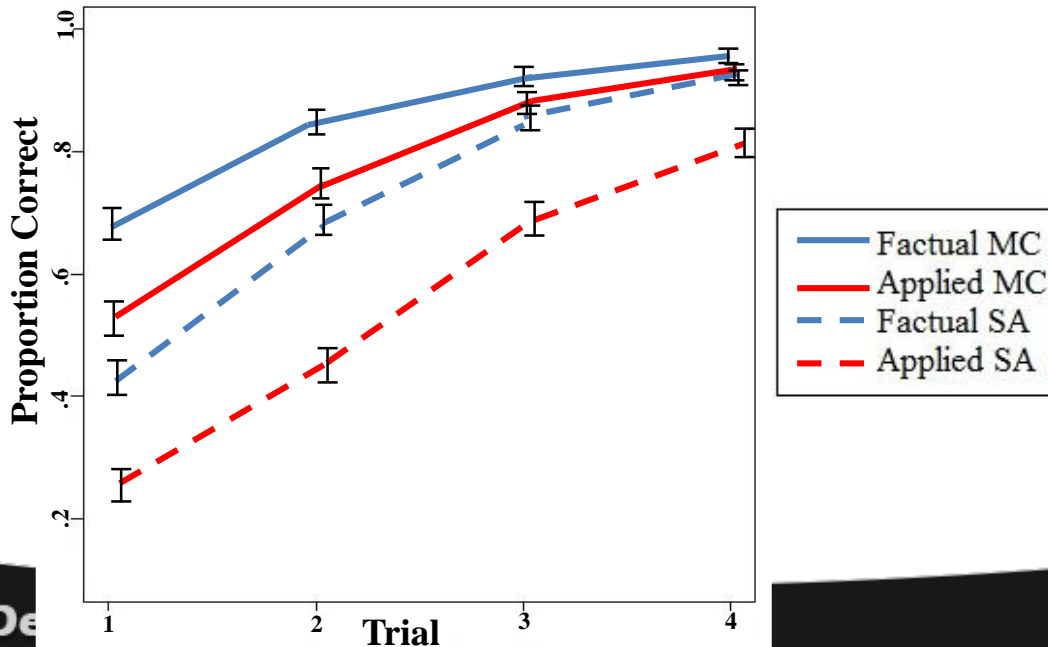
Revisiting our Questions

- Can we increase the difficulty of items through either format (multiple choice vs. short answer) and/or through depth (factual vs. applied)? ✔ Yes ☐ No

- Does retrieval need to be successful in order to be beneficial? ✔ Yes ✔ No

  successful retrieval of

  ⬇

- Do participants benefit from more difficult items? ✔ Yes ☐ No
  - If so, is the cause of the increase in difficulty an important factor (i.e., depth or <u>format</u>)? ✔ Yes ☐ No

INSTITUTE for INTELLIGENT SYSTEMS

Department of Psychology

THE UNIVERSITY OF MEMPHIS

Implications

- Scheduling practice based on our results
  - Early on, use multiple choice to increase chance of success   or factual

  - Transition into short answer to get the most out of practice   or applied

INSTITUTE for INTELLIGENT SYSTEMS

THE UNIVERSITY OF MEMPHIS

Thank you!

Any Questions

Website: **https://sites.google.com/site/jaclynmaass**
Email: **Jaclyn.maass@gmail.com**

INSTITUTE *for* INTELLIGENT SYSTEMS

THE UNIVERSITY OF
MEMPHIS