

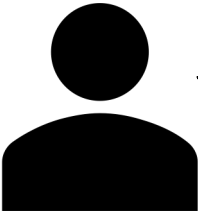
# How effective is interleaved practice for teaching statistics?

Paulina N. Silva & Barbara W. Sarnecka  
Department of Cognitive Sciences  
University of California, Irvine  
USCOTS 2021

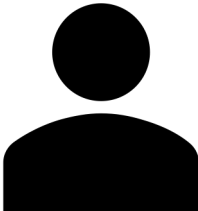
Stats is hard.

# Why is stats hard?

- Conceptual (mis)understanding
- Conceptual (mis)application



“I know this fact, but I don’t know how to use it...”



“ $p$ -values tell us the probability that the experiment worked!”

How can we improve students'  
conceptual understanding?

# Interleaving to learn concepts

- Used to successfully teach concepts and categories
- More effective than blocking

Interleaving:

A B C A B C A B C

Blocking:

A A A B B B C C C

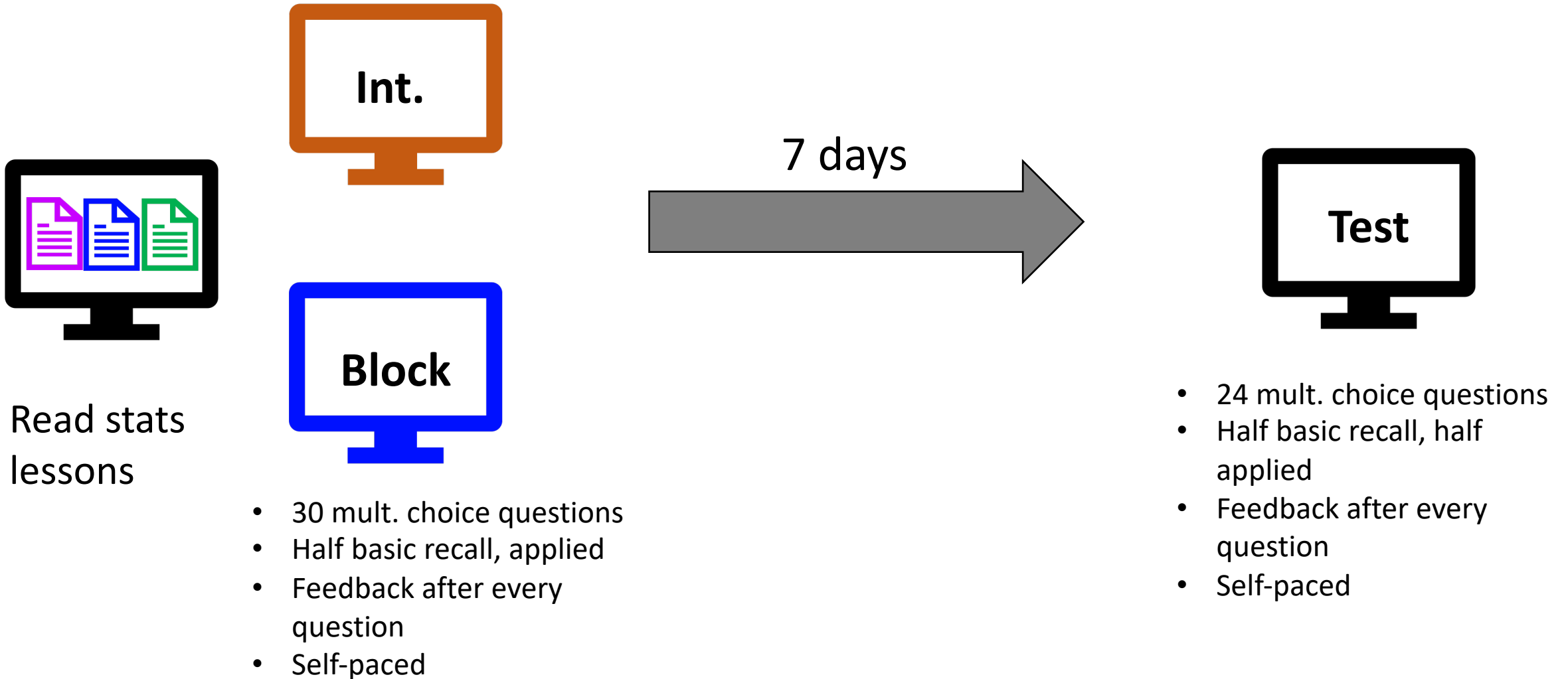
# Hypotheses

- Interleaved practice will result in:
  1. Better overall performance at test
  2. Better performance on applied problems at test

# Participants

- Final sample:
  - 64 participants (3 removed)
  - 90.6% female
  - Avg. age=20.63 y (SD=4.39)
  - First gen=59.36%
  - URM=59.36%
  - Modal stats experience: 2-3 courses
- Recruited through human subjects research pool
- 0.5 unit credit per 30 min. of participation; max credit was 1 unit

# Materials and Procedure





# Materials and Procedure

**Basic recall:** “What does central tendency measure?”

- a) How similar or dissimilar a set of scores are
- b) The typical score in a set of scores**
- c) Whether one group is different from another
- d) Skewness of data

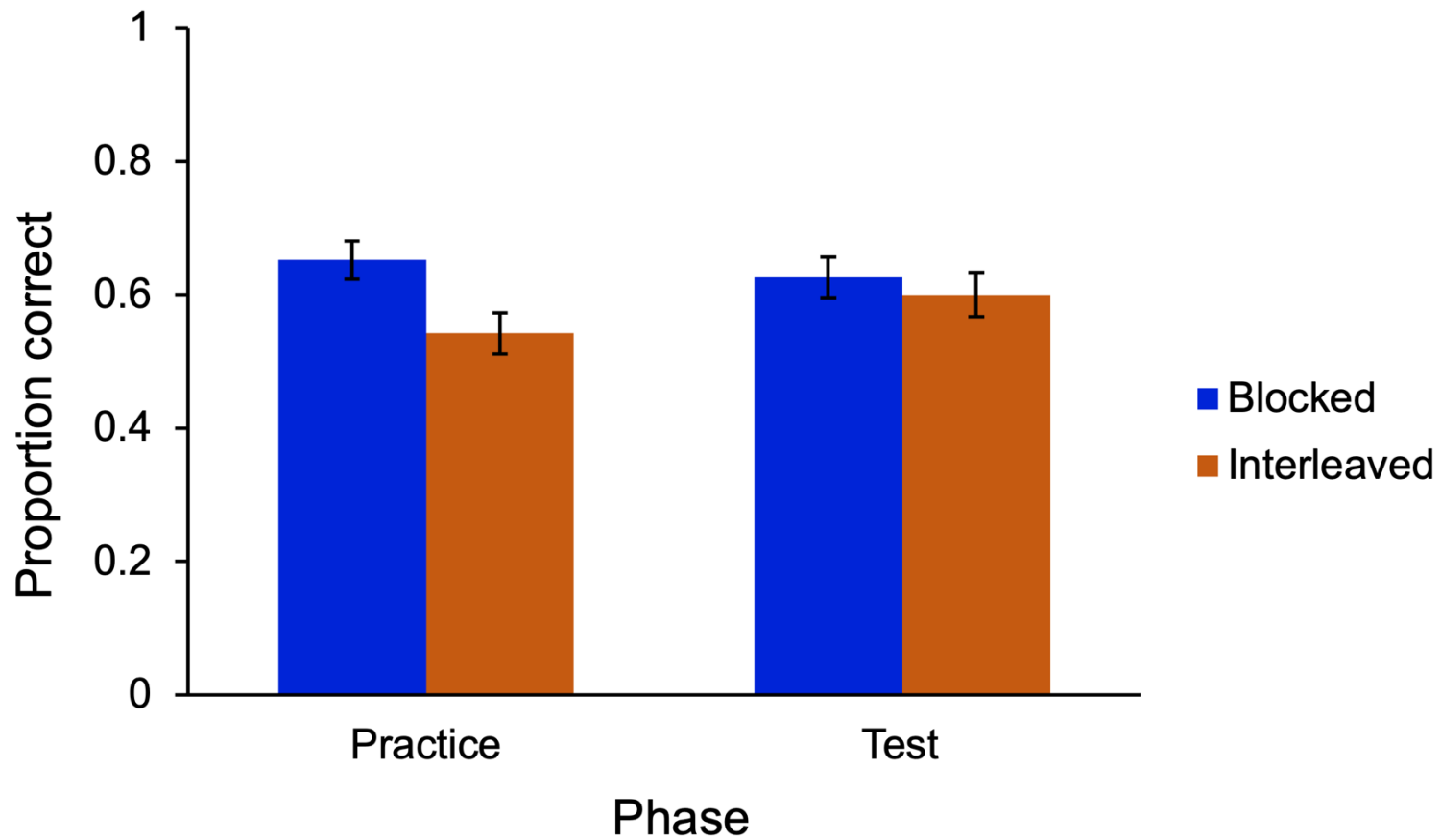
# Materials and Procedure

**Applied:** “*Introduction to Psychology* students took an exam. Here are the scores from nine students: 40%, 50%, 70%, 80%, 55%, 65%, 50%, 60%, 70%. Which of the following are three different ways to measure the typical score of these students?”

- a) variance, mean, and mode
- b) density, median, and average
- c) mode, mean, and median**
- d) mean, frequency, and variance

# Results

Total Proportion Correct



Controlled for previous stats experience and stats anxiety

- Practice:  $F(1, 59)=2.89, p=.10$
- Phase:  $F(1, 59)=.67, p=.42$
- **Practice\*phase:  $F(1, 59)=9.76, p=.003, \eta_p^2=.14$**

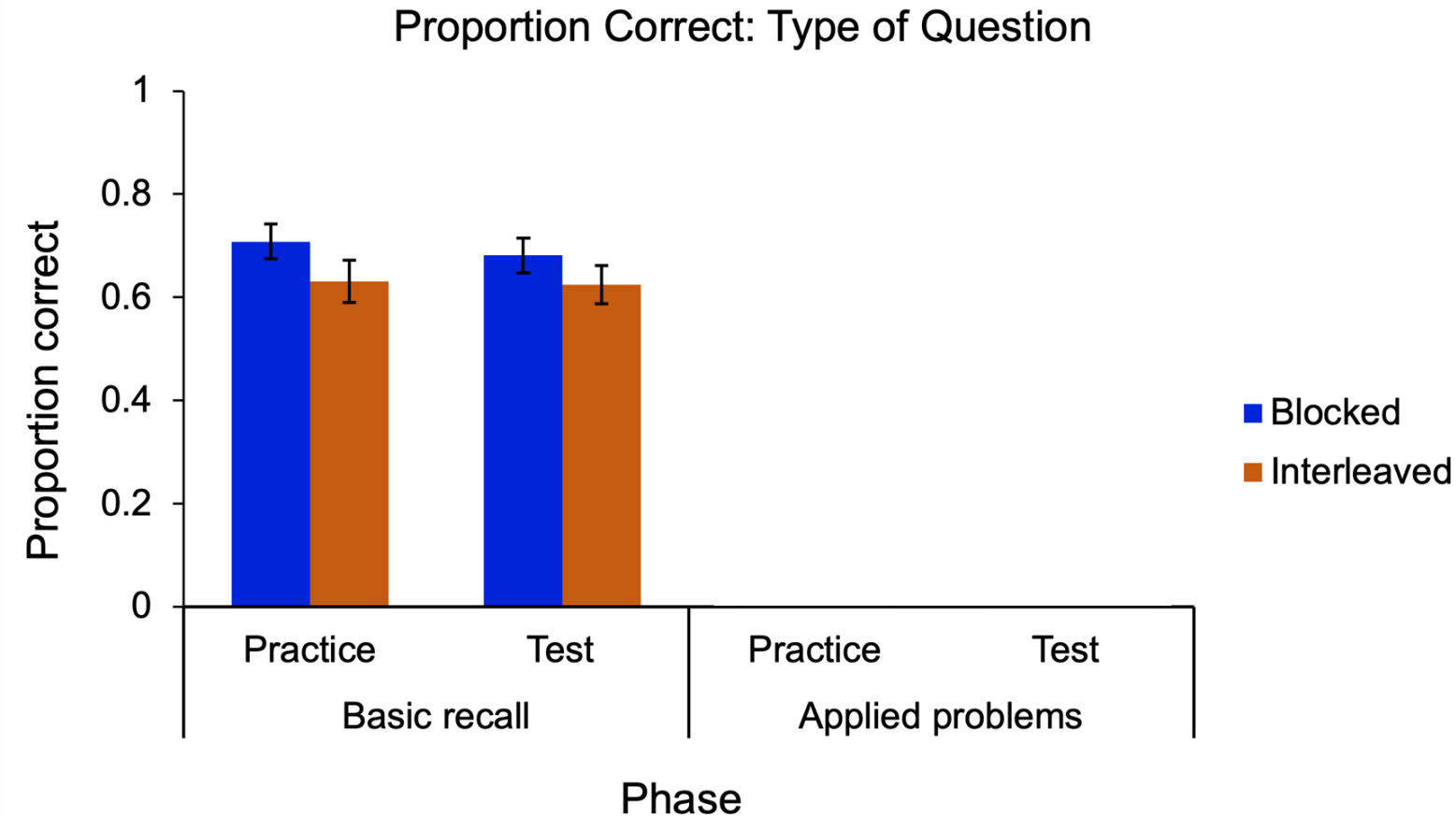
# Results

Controlled for previous stats experience and stats anxiety

## Basic recall:

- Practice:  $F(1, 59)=1.96, p=.17$
- Phase:  $F(1, 59)=.21, p=.65$
- Practice\*Phase:  $F(1, 59)=.37, p=.55$

## Applied problems:



Interleaving didn't seem to help

# What might have happened?

- Task potentially too hard



# What might have happened?

- Existing contexts interleaving has been studied in may not generalize to stats



# Future directions

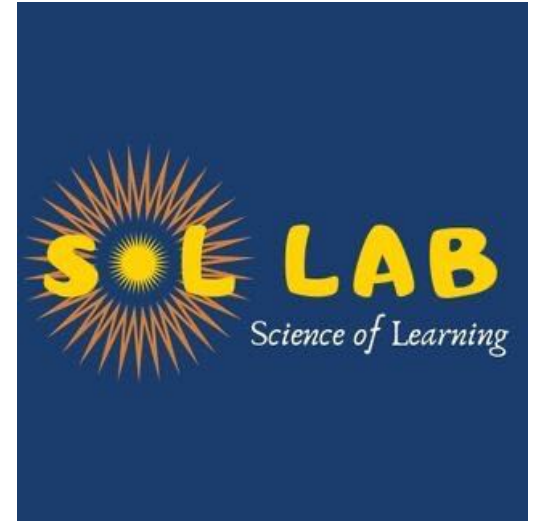
- In what contexts is interleaving effective for teaching *complex, abstract* concepts?



# Thank you!



Thank you to Barbara and all the members of the Sarnecka lab, past and current, for your support and feedback



Thank you to Lindsey Richland and the SOL lab for all your feedback and discussions about this project



@pnsilva\_



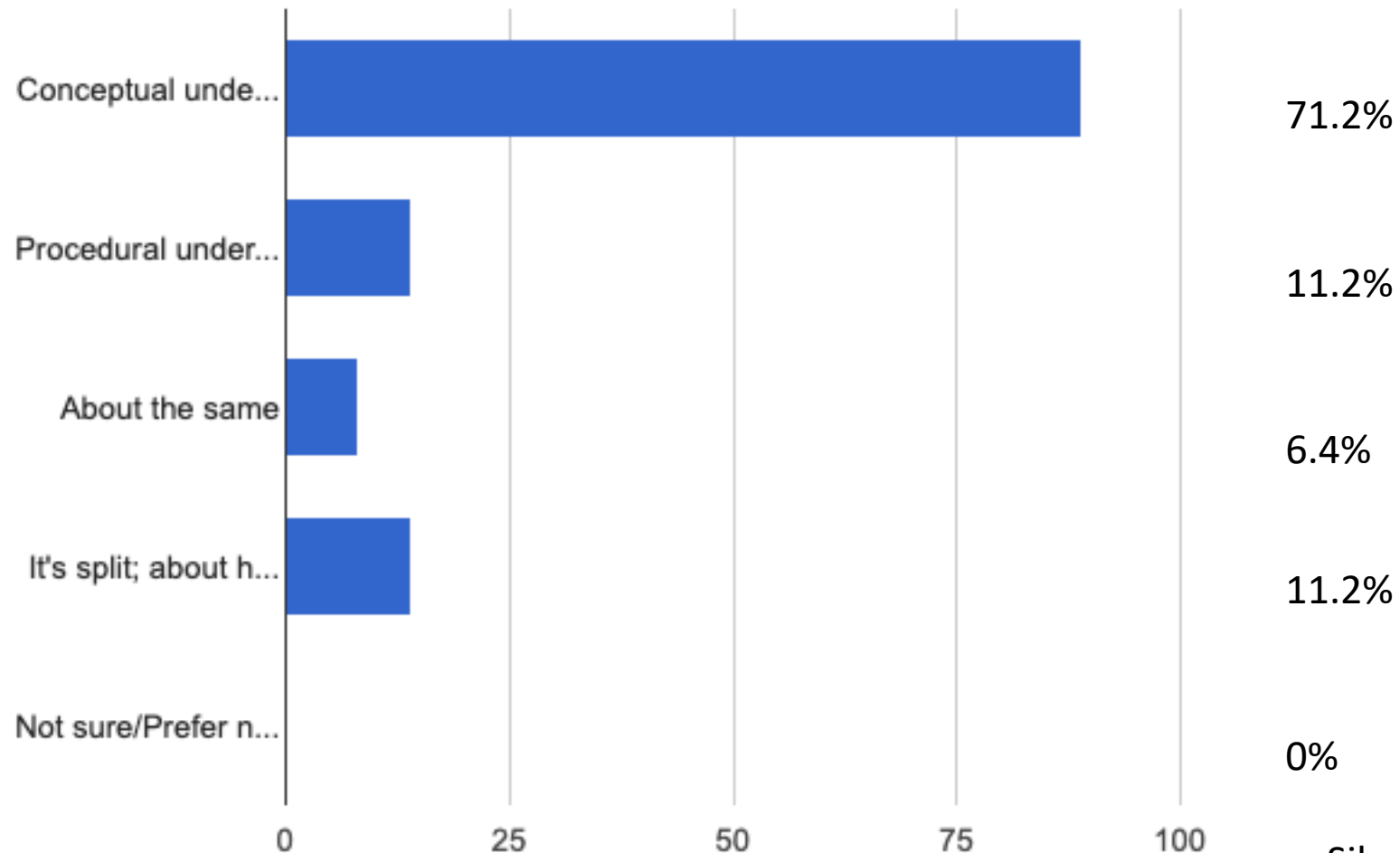
pnsilva@uci.edu

# Supplementary Slides

Additional background, secondary findings, and Bayesian analyses

# Survey of stats instructors (in progress)

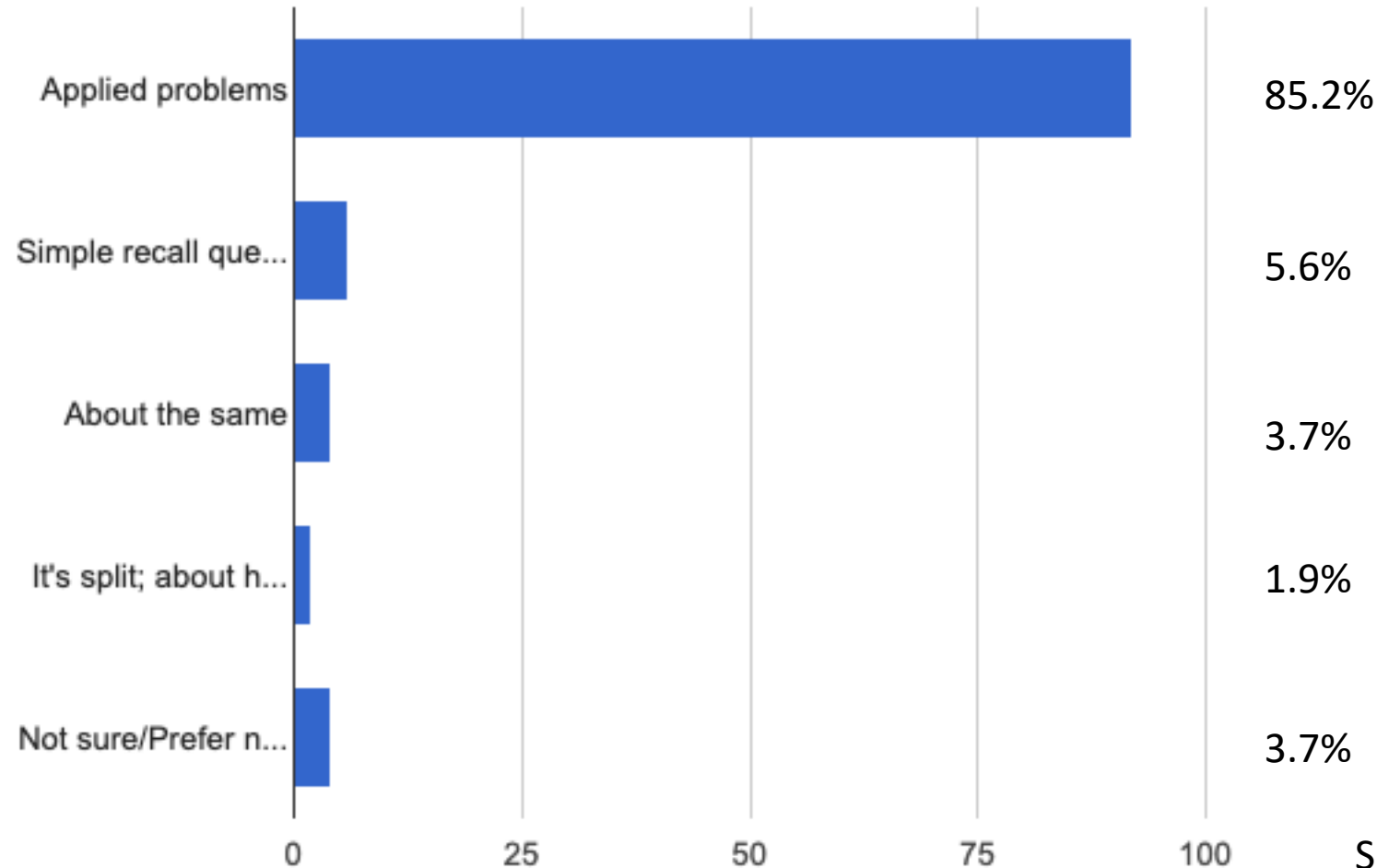
Which is harder: conceptual or procedural understanding? ( $n=125$ )



Silva et al., [data collection in progress] (2021)

# Survey of stats instructors (in progress)

Which is harder: simple recall questions or applied problems? ( $n=108$ )



# Results: Secondary findings

- **Total proportion correct**
  - **Higher stats anxiety predicted lower scores:**
    - $F(1, 59)=10.90, p=.002, \eta_p^2=.16$
- **Basic recall and applied problems, separately**
  - **Higher stats anxiety predicted lower scores:**
    - $F_{basic}(1, 59)=9.25, p=.004, \eta_p^2=.14$
    - $F_{app}(1, 59)=10.69, p=.002, \eta_p^2=.15$

# Results: Bayesian analyses

## Total proportion correct

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.050	0.007	0.138	1.000	
Phase + Practice + attitude_score + Phase*Practice	0.050	0.306	8.391	42.598	4.879
Phase + Practice + previous_stats + attitude_score + Phase*Practice	0.050	0.228	5.617	31.730	13.243
Practice + attitude_score	0.050	0.104	2.206	14.468	2.844
attitude_score	0.050	0.086	1.779	11.904	1.291
Practice + previous_stats + attitude_score	0.050	0.068	1.382	9.427	4.733
previous_stats + attitude_score	0.050	0.057	1.143	7.891	3.459
Phase + Practice + attitude_score	0.050	0.027	0.525	3.740	2.292
Phase + attitude score	0.050	0.025	0.488	3.485	5.175
Phase + Practice + Phase*Practice	0.050	0.022	0.425	3.045	3.956
Phase + Practice + previous_stats + attitude_score	0.050	0.018	0.353	2.534	14.663
Phase + previous_stats + attitude_score	0.050	0.016	0.307	2.213	5.014
Phase + Practice + previous_stats + Phase*Practice	0.050	0.014	0.277	1.999	16.181
Practice	0.050	0.007	0.137	0.994	1.478
previous_stats	0.050	0.004	0.079	0.573	1.501
Practice + previous_stats	0.050	0.004	0.076	0.554	3.794
Phase + Practice	0.050	0.002	0.041	0.298	2.779
Phase	0.050	0.002	0.038	0.275	1.339
Phase + previous_stats	0.050	0.001	0.024	0.175	8.694
Phase + Practice + previous_stats	0.050	0.001	0.020	0.149	2.475

Note. All models include subject

# Results: Bayesian analyses

## Proportion correct: Basic recall

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.050	0.032	0.623	1.000	
attitude_score	0.050	0.238	5.946	7.510	1.653
Practice + attitude_score	0.050	0.193	4.543	6.080	2.688
previous_stats + attitude_score	0.050	0.149	3.320	4.687	0.890
previous_stats + Practice + attitude_score	0.050	0.113	2.420	3.560	10.058
Phase + attitude_score	0.050	0.057	1.146	1.792	2.226
Phase + Practice + attitude_score	0.050	0.045	0.890	1.410	6.183
Phase + previous_stats + attitude_score	0.050	0.035	0.686	1.098	9.152
Phase + previous_stats + Practice + attitude_score	0.050	0.029	0.570	0.918	6.975
previous_stats	0.050	0.024	0.459	0.744	0.773
Practice	0.050	0.023	0.450	0.730	0.793
Phase + Practice + attitude_score + Phase * Practice	0.050	0.014	0.271	0.444	4.466
previous_stats + Practice	0.050	0.012	0.231	0.379	2.626
Phase + previous_stats + Practice + attitude_score + Phase * Practice	0.050	0.009	0.179	0.293	4.733
Phase	0.050	0.008	0.156	0.256	0.807
Phase + Practice	0.050	0.008	0.145	0.239	18.546
Phase + previous_stats	0.050	0.006	0.117	0.193	2.481
Phase + previous_stats + Practice	0.050	0.003	0.057	0.094	1.589
Phase + Practice + Phase * Practice	0.050	0.002	0.033	0.055	2.131
Phase + previous_stats + Practice + Phase * Practice	0.050	8.961e -4	0.017	0.028	3.556

Note. All models include subject



# Results: Bayesian analyses

## Proportion correct: Applied problems

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.050	0.025	0.481	1.000	
attitude_score	0.050	0.403	12.841	16.319	1.643
previous_stats + attitude_score	0.050	0.172	3.935	6.943	0.862
Practice + attitude_score	0.050	0.120	2.601	4.872	1.571
Phase + attitude_score	0.050	0.091	1.902	3.683	1.783
previous_stats + Practice + attitude_score	0.050	0.058	1.169	2.346	6.680
Phase + previous_stats + attitude_score	0.050	0.040	0.800	1.635	1.210
Phase + Practice + attitude_score	0.050	0.028	0.554	1.147	2.587
Phase + previous_stats + Practice + attitude_score	0.050	0.014	0.264	0.554	5.740
previous_stats	0.050	0.013	0.258	0.541	1.874
Phase + Practice + attitude_score + Phase * Practice	0.050	0.008	0.160	0.337	2.508
Practice	0.050	0.008	0.154	0.325	0.653
Phase	0.050	0.006	0.110	0.232	0.986
Phase + previous_stats + Practice + attitude_score + Phase * Practice	0.050	0.004	0.069	0.146	2.769
Phase + previous_stats	0.050	0.003	0.063	0.133	2.324
previous_stats + Practice	0.050	0.003	0.055	0.118	1.000
Phase + Practice	0.050	0.002	0.036	0.077	1.586
Phase + previous_stats + Practice	0.050	6.845e -4	0.013	0.028	3.183
Phase + Practice + Phase * Practice	0.050	5.260e -4	0.010	0.021	1.632
Phase + previous_stats + Practice + Phase * Practice	0.050	2.096e -4	0.004	0.008	5.472

Note. All models include subject