**Supplemental Results**

**Risk of Bias Moderator Analyses**

***Allocation concealment:*** *Q-M* = .89, *df* = 2, *p* = .64

***Incomplete outcome reporting:*** *Q-M =* 3.0, *df* =1, *p* = .08

**Selective outcome reporting:** Q-M = 1.72, *df* = 2, *p* = .42

***Sequence generation:*** *Q-M =* 3.5, *df* = 1, *p* = .06

***Blinding of outcome assessors:*** *Q- M=* 1.3, *df* = 2, *p* = .51

***Blinding of participants and personnel:*** *Q-M* = 2.18, *df* = 2, *p* = .34

***Other sources of bias:*** *Q-M* = 0.45, *df* = 1, *p* = .50

**Sensitivity Analyses**

**Outlier Removed**

***Multilevel Estimates***

**Acquisition:** *g* **=** .23, *p* = .10, 95% CI [-.04, .51]

**Immediate retention:** *g* =.12, *p* = .43, 95% CI [-.18, .41]

**Delayed retention:** *g* = .19, *p* = .10, 95% CI [-.04, .42]

***Categorical*** ***Moderator Analyses***

**Age:** *Q-M = 5.42,* df *=* 6, *p* = .49

**Skill:** *Q-M* = 1.1, *df* = 3, *p* = .78

**Faded**: *Q-M* = 2.17, *df* = 4, *p* =.70

**Yoked:** *Q-M* = 6.27, *df* = 4, *p* = .18

**Feedback:** *Q-M* = 7.64, *df* = 8, *p* = .47

**Measure:** *Q-M* = 30.65, *df* = 22, *p* = .10

***Measure without test interaction*:** *Q-M* = 9.07, *df* = 8, *pi = .34*

**Bandwidth:** *Q-M* = 3.35, *df* = 5, *p* = .65

***Mete-regression Analyses***

**Trials:** *Q-M* = 5.80, *df* = 3, *p* = .12

**Days:** *Q-M* = 2.72, *df* = 3, *p* = .44

**Frequency (overall analysis):** *Q-M* = 1.99, *df* = 3, *p* = .57

**Immediate retention interval:** *Q-M* = .087, *df* = 2, *p* = .65

***Test Time Moderators***

**Immediate retention vs. delayed retention:** *Q-M* = .24, *df* = 1, *p* = .63.

**Cluster Robust Inference Methods**

***Cluster Robust Multilevel Estimates***

**Acquisition:** *g* = .19, *p* = .20, 95% CI [-.11, .50]

**Immediate retention:** *g* = .14, *p* = .93, 95% CI [-.29, .31]

**Delayed retention** *g* = .19, *p* = .15, 95% Ci [-.07, .46]

***Correlated and Hierarchical Effects (CHE) Model Estimates with Approximate V Matrix (r = .7)***

**Acquisition:** *g* = .19, *p* = .22, 95% CI [-.13, .51]

**Immediate retention:** *g* = .002, *p* = .99, 95% CI [-.34, .34]

**Delayed retention:** *g* = .20, *p* = .13, 95% CI [-.07, .48]

***Cluster Robust Multilevel Estimates with Outlier Removed***

**Acquisition:** *g* = .23, *p* = .11, 95% CI [-.06, .52]

**Immediate retention:** *g* = .12, *p* = .34, 95% CI [-.13, .37]

**Delayed retention** *g* = .19, *p* = .14, 95% Ci [-.07, .45]

***CHE Model Estimates with Approximate V Matrix (r = .7) and Outlier Removed***

**Acquisition:** *g* = .23, *p* = .13, 95% CI [-.08, .53]

**Immediate retention:** *g* = .12, *p* = .36, 95% CI [-.15, .38]

**Delayed retention:** *g* = .20, *p* = .13, 95% CI [-.06, .46]

**Four Level Model: Measure Nested in Test Nested in Experiment**

***Multilevel Estimates***

**Acquisition:** *g* **=** .15, *p* = .13, 95% CI [-.04, .34]

**Immediate retention:** *g* =.07, *p* = .56, 95% CI [-.15, .29]

**Delayed retention:** *g* = .18, *p* = .051, 95% CI [-.001, .36]

***Four Level Model: Outlier Removed***

**Multilevel Estimates**

***Acquisition:*** *g* **=** .18, *p* = .07, 95% CI [-.01, .37]

***Immediate retention:*** *g* =.12, *p* = .27, 95% CI [-.09, .34]

***Delayed retention:*** *g* = .18, *p* = .051, 95% CI [-.001, .36]

***Moderator Analyses***

**Age:** *Q-M* = 7.81, *df* = 6, *p* = .25

**Age (outlier removed):** *Q-M* = 6.06, *df* = 6, *p* = .42

**Skill:** *Q-M* = 1.75, *df* = 3, *p* = .63

**Skill (outlier removed):** *Q-M* = .98, *df* = 3, *p* = .81

**Task:** *Q-M* = 19.47, *df* = 8, *p* = .01

**Task (Drews et al. 2021 removed):** *Q-M* = 9.4, *df* = 8, *p* = .31

**Bandwidth:** *Q-M* = 1.74, *df* = 5, *p* = .88

**Bandwidth (outlier removed):** *Q-M* = 1.19, *df* = 5, *p* = .94

**Faded:** *Q-M* = 3.03, *df* = 5, *p* = .70

**Faded (outlier removed):** *Q-M* = 2.36, *df* = 5, *p* = .80

**Yoked:** *Q-M* = 6.49, *df* = 4, *p* = .17

**Yoked (outlier removed):** *Q-M* = 6.22, *df* = 4, *p* = .18

**Feedback:** *Q-M* = 11.66, *df* = 16, *p* = .77

**Feedback (outlier removed):** *Q-M* = 8.53, *df* = 16, *p* = .93

**Feedback (interaction removed):** *Q-M* = 4.15, *df* = 6, *p* = .66

**Measure:** *Q-M* = 16.17, *df* = 28, *p* = .96

**Measure (outlier removed):** *Q-M* = 14.49, *df* = 28, *p* = .98.

**Measure (interaction removed):** *Q-M* = 5.24, *df* = 10, *p* = .87.

***Do Measures Selected as Primary Differ from Secondary Measures***

**Full sample:** *Q-M* = 1.85, *df* = 5, *p* = .87

**Outlier removed:** *Q-M* = 1.08, *df* = 5, *p* = .96

***Meta-regression Analyses***

**Trials:** *Q-M* = 9.83, *df* = 5, *p* = .08

**Trials (outlier removed):** *Q-M* = 9.92, *df* = 5, *p* = .08

**Days:** *Q-M* = 7.75, *df* = 5, *p* = .17

**Days (outliers removed):** *Q-M* = 7.70, *df* = 5, *p* = .17

**Days:** *Q-M* = 7.75, *df* = 5, *p* = .17

**Univariate Analysis of Transfer Test Data**

***Estimate:*** *g* = .15, *p =* .45, 95% CI [-.24, .55]

***Heterogeneity:*** *Q* = 68.47, *df* = 14, *p* < .0001. τ2= .57

***Estimate (outlier removed):*** *g* = .05, *p =* .83, 95% CI [-.38, .48]

***Heterogeneity (outliers removed):*** *Q* = 55.19, *df* = 13, *p* < .0001. τ2= .42