**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

**Feedback Collapsed: Spatial error, temporal error, variable error, movement time, form, and other.**

**Feedback:** *Q-M* = 11.41, *df* = 14, *p* = .78

**Feedback (outlier removed):** *Q-M* = 8.71, *df* = 14, *p* = .92

**Feedback (interaction removed):** *Q-M* = 2.21, *df* = 5, *p* = .82

**Feedback (delayed retention only):** *Q-M* = 2.07, *df* = 5, *p* = .84

**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

**Moderator Analyses of Spatial Error Subset and Delayed Retention Time Point**

***Age:*** *Q-M* = 1.72, *df* = 1, *p* = .19

***Skill:*** *Q-M* = .61, *df* = 1, *p* = .44

***Task:*** *Q-M* = 1.77, *df* = 2, *p* = .41

***Faded:*** *Q-M* = .72, *df* = 1, *p* = .40

***Yoked:*** *Q-M* = .29, *df* = 1, *p* = .59

***Feedback:*** *Q-M* = 5.08, *df* = 4, *p* = .28

***Trials:*** *Q-M* < .001, *df* = 1, *p* = .98

***Days:*** *Q-M* = .002, *df* = 1, *p* = .96

***Frequency:*** *Q-M* = 1.14, *df* = 1, *p* = .29

**Simulation Study**

1. # # # # # # # # # # # # # # # # # # # # # # # # # # #
2. #Meta-analysis of dependent effect sizes simulation #
3. # # # # # # # # # # # # # # # # # # # # # # # # # # #
4. library(MBESS)
5. library(metafor)
6. library(faux)
7. library(tidyverse)
8. set.seed(1000)


12. # individual meta-analysis simulation function
14. onemeta <- function(nSims, cxa,exa, cxb, exb, rho, sample\_lb, sample\_ub){
15. esd\_1 <-numeric(nSims) #empty container for all simulated ES
16. esd\_2 <-numeric(nSims) #empty container for all simulated ES
17. SSn1 <-numeric(nSims) #empty container for random sample sizes group 1
18. SSn2 <-numeric(nSims) #empty container for random sample sizes group 2
19. for(i in 1:nSims){ #for each simulated experiment
20. SampleSize<-sample(sample\_lb:sample\_ub, 1) #randomly draw a sample between lb and ub
21. ge <-rnorm\_multi(n = SampleSize, 2, 0, 1, r = rho, varnames = c("t1", "t2")) #sample from a multivariate normal
22. gc <-rnorm\_multi(n = SampleSize, 2, 0, 1, r = rho, varnames = c("t1", "t2")) #sample from a multivariate normal
23. gc$t1 <- gc$t1 + cxa # add effect to time 1 for control
24. ge$t1 <- ge$t1 + exa # add effect to time 1 for experimental
25. gc$t2 <- gc$t2 + cxb # add effect to time 2 for experimental
26. ge$t2 <- ge$t2 + exb # add effect to time 2 for experimental
27. SSn1[i]<-SampleSize #save sample size group 1
28. SSn2[i]<-SampleSize #save sample size group 2
29. esd\_1[i]<-smd(Mean.1= mean(ge$t1), Mean.2=mean(gc$t1), s.1=sd(ge$t1), s.2=sd(gc$t1), n.1=SampleSize, n.2=SampleSize, Unbiased=TRUE) #Use MBESS to calc Hedges g
30. esd\_2[i] <-smd(Mean.1= mean(ge$t2), Mean.2=mean(gc$t2), s.1=sd(ge$t2), s.2=sd(gc$t2), n.1=SampleSize, n.2=SampleSize, Unbiased=TRUE)}
31. #Insert effect sizes and sample sizes
32. n1<-c(SSn1)
33. n2<-c(SSn2)
34. J<-1-3/(4\*( SSn1+ SSn2-2)-1) #correction for bias
35. esdv\_1 <-(((SSn1+SSn2)/(SSn1\*SSn2))+(esd\_1^2/(2\*(SSn1+SSn2))))\*J^2
36. esdv\_2 <-(((SSn1+SSn2)/(SSn1\*SSn2))+(esd\_2^2/(2\*(SSn1+SSn2))))\*J^2
37. dat <- as.data.frame(cbind(esd\_1, esdv\_1, esd\_2, esdv\_2, n1, n2))
38. dat$id <- seq.int(nrow(dat))
39. dat<-dat %>%
40. pivot\_longer(
41. cols = !c(n1,n2,id),
42. names\_to = c(".value", "num"),
43. names\_sep = "\_")
44. return(as.data.frame(dat))
46. }


50. # simulation study ## multilevel models
52. iters <- 100 #number of simulated meta-analyses
53. mpval <- rep(NA, iters) #empty container for moderator p-values
54. difb <- rep(NA, iters) #empty container for moderator beta values
55. nSims <- 32 #number of experiments per meta-analysis
56. cxa <- 0 #intervention effect for control group at first time point
57. exa <- 0 #intervention effect for experimental group at first time point
58. cxb <- 0 #intervention effect for control group at second time point
59. exb <- 0 #intervention effect for experimental group at second time point
60. rho <- .5 #correlation between time points in the population
61. sample\_lb <- 10 #lower bound for number of participants per experiment
62. sample\_ub <- 30 #upper bound for number of participants per experiment



67. for (i in 1:iters){
68. dat <-onemeta(nSims = nSims, cxa = cxa, exa = exa, cxb = cxb, exb = exb, rho = rho, sample\_lb = sample\_lb, sample\_ub = sample\_ub)
69. tryCatch(
70. {
71. res <- rma.mv(esd, esdv, mods = ~factor(num), random = ~1|id/num, data = dat) #fit multilevel model with time points nested in experiments
72. mpval[i] <- res$QMp
73. difb[i] <- res$b[2]},
74. error=function(error\_message) {
75. message(error\_message)
76. return(NA) #when the model fails to converge, store as NA rather than stopping the loop
77. }
78. )
79. }
81. mean(mpval <= .05, na.rm = TRUE) # mean rejection rate with alpha = .05
82. mean(difb, na.rm = TRUE) # mean estimate of difference in intervention effect between time points





89. # simulation study ## cluster robust methods
91. iters <- 100 #number of simulated meta-analyses
92. mpval <- rep(NA, iters) #empty container for moderator p-values
93. difb <- rep(NA, iters) #empty container for moderator beta values
94. nSims <- 32 #number of experiments per meta-analysis
95. cxa <- 0 #intervention effect for control group at first time point
96. exa <- 0 #intervention effect for experimental group at first time point
97. cxb <- 0 #intervention effect for control group at second time point
98. exb <- 0 #intervention effect for experimental group at second time point
99. rho <- .5 #correlation between time points in the population
100. sample\_lb <- 10 #lower bound for number of participants per experiment
101. sample\_ub <- 30 #upper bound for number of participants per experiment


105. for (i in 1:iters){
106. dat <-onemeta(nSims = nSims, cxa = cxa, exa = exa, cxb = cxb, exb = exb, rho = rho, sample\_lb = sample\_lb, sample\_ub = sample\_ub)
107. tryCatch(
108. {
109. res <- rma.mv(esd, esdv, mods = ~factor(num), random = ~1|id/num, data = dat)
110. rob <-robust(res, cluster = dat$id) #use cluster robust inference methods
111. mpval[i] <- rob$QMp
112. difb[i] <- rob$b[2]},
113. error=function(error\_message) {
114. message(error\_message)
115. return(NA)
116. }
117. )
118. }
120. mean(mpval <= .05, na.rm = TRUE) # mean rejection rate with alpha = .05
121. mean(difb, na.rm = TRUE) # mean estimate of difference in intervention effect between time points

**To conduct a simulation, adjust the following settings:**

**Iters: Number of meta-analyses to simulate (warning, the simulation can take a long time to run if this number is large)**

**nSims: Number of studies per meta-analysis.**

**cxa: The size of the effect to add to the first time point for the control group. Effects are scaled to Cohen’s *d*.**

**exa: The size of the effect to add to the first time point for the experimental group.**

**cxb: The size of the effect to add to the second time point for the control group.**

**exb: The size of the effect to add to the second time point for the experimental group.**

**rho: The correlation between time points in the population, prior to adding intervention effects.**

**sample\_lb: The lower bound for number of participants per group in an experiment included in a simulated meta-analysis. Groups will be equal in size for each experiment.**

**sample\_ub: The upper bound for number of participants per group.**

**This simulation can be extended to return additional results from meta-analysis models by adding containers above the simulation and specifying the output to save within the simulation.**