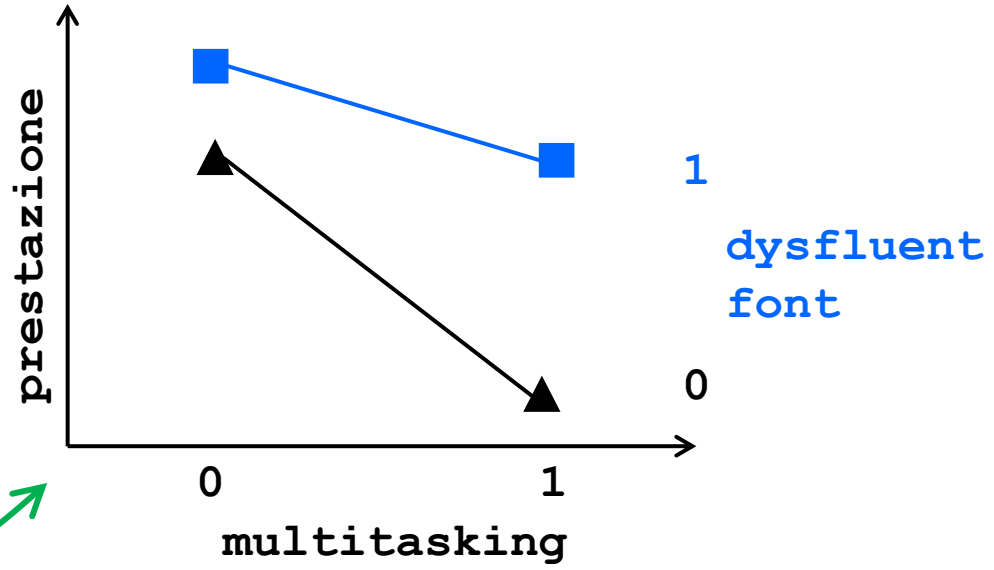
Quale ipotesi?

Cosa dobbiamo capire (e poi dare dei numeri!)

- Qual è l'effetto del **MULTITASKING**? Supponiamo ***negativo***: compiti distraenti durante il compito principale riducono la performance.
- Qual è l'effetto della **DYSFLUENCY**? Supponiamo ***positivo***: *font* di difficile lettura costringe a maggiore dedizione al testo e migliora la comprensione (?)
- **Interazione MULTITASKING x DYSFLUENCY**? I due effetti si limitano a sommarsi quando co-presenti? oppure la DYSFLUENCY porta a una «supercompensazione» del MULTITASKING, es. annullandone interamente l'effetto?



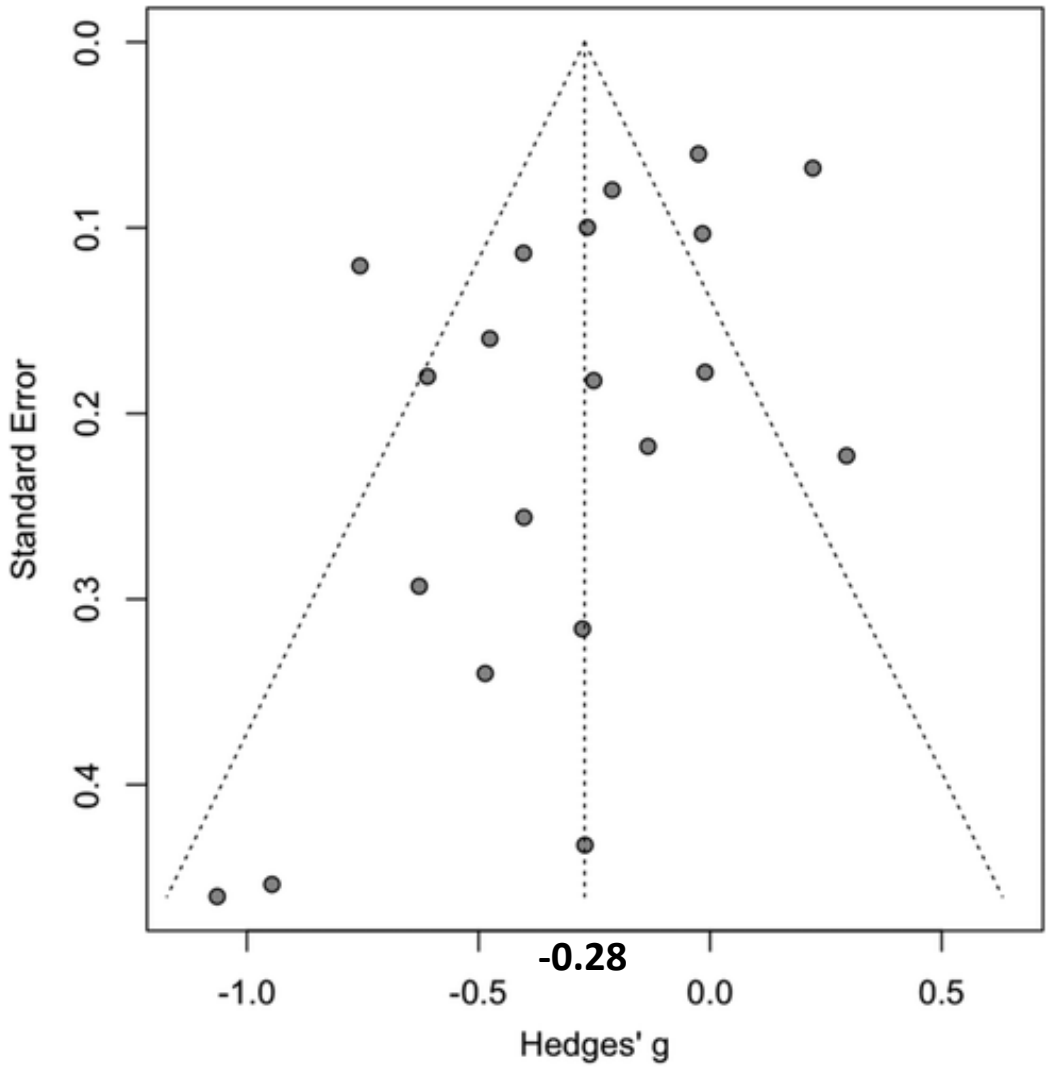
ipotesi con *multitasking* che ha impatto decisamente minore se *dysfluent text* è co-presente

MULTITASKING – Main effect

Clinton-Lisell, V. (2021). Stop multitasking and just read: Meta-analyses of multitasking's effects on reading performance and reading time. *Journal of Research in Reading, 44*(4), 787-816.

RQ1: Reading comprehension
(…) Based on the RVE analyses of 31 effect sizes within 20 independent studies and a total of 1,686 participants, readers who multitasked had worse reading comprehension than readers who did not multitask, **g = −0.28**, standard error (SE) = 0.07, 95% confidence interval (CI) [−0.43, −0.12], p = .002 (…) Publication bias was tested using two methods: a funnel plot and Egger's test of the intercept (…) Egger's test of the intercept indicated a reliable difference from zero, β = −2.03, (−3.60, −0.46), t = −2.39, p = .03. Taken together, publication bias is quite possible, although the inclusion of grey literature, such as the undergraduate theses, in these findings would prevent publication bias from being a substantial concern (…) The overall effect size with the removal of these outliers (18 studies with 26 effect sizes) indicated similar findings with their inclusion, g = −0.26, SE = 0.06, 95% CI [−0.39, −0.12], p = .001, indicating readers in multitasking conditions performed less well on comprehension assessments

MA metodi che mettono in relazione effetto e sua varianza (come Egger's test e Funnel plot) inflazionano falsi positivi del publication bias per *Standardized Mean Differences*, quindi tengo l'effetto com'è



DYSFLUENCY – Main effect

- Articolo di Dieman-Yauman et al. (2011). Cognition. 118, 111–115.
Primo studio in laboratorio: non è riportato. Secondo studio in classe – memory: **Cohen's d = 0.45**

- Articolo di Halin et al. (2014). Journal of Applied Research in memory and Cognition, 3, 31–36.
Memory for text: no significant main effect of task difficulty, $F(1, 31) = 0.12$, $p = .729$, **ηp2 = .004**

- Articolo di Eitel et al. (2014). Applied Cognitive Psychology, 28, 488–501.
Only in Exp. 1 effect. Retention: no effect.
Transfer: main effect of text legibility, $F(1, 79) = 5.62$, $p = 0.02$, **ηp2 = 0.07**

- Articolo di Halin et al. (2016). Frontiers in Psychology, 7.
Text memory: no main effect of font condition, $F(1,54) = 0.49$, $p = 0.489$, **ηp2 = 0.009**

- Articolo di Lai e Zhang (2021). Frontiers in Psychology, 12. 755804.
Retention: no effect. Transfer: main effect for fluency, $F(1,60) = 4.481$, $p = 0.038$, **ηp2 = 0.069**

- Articolo di Hao & Conway (2022). Memory & Cognition, 50, 852–863.
Comprehension: The main effect of perceptual disfluency was significant, $F(1, 120) = 5.42$, $p = 0.022$, **ηp2 = .04**

DYSFLUENCY – Main effect

Study	N	eff	vi
Dieman-Yauman et al. (2011)	222	0.4441	0.0183
Eitel et al. (2014)	42	0.2857	0.0925
Eitel et al. (2014)	42	0.5487	0.0951
Halin et al. (2014)	32	0.1267	0.1189
Halin et al. (2016)	56	0.1906	0.0697
Lai & Zhang (2021)	32	0.2225	0.1194
Lai & Zhang (2021)	32	0.9741	0.1334
Hao & Conway (2022)	126	0.4082	0.0320

DYSFLUENCY – Main effect

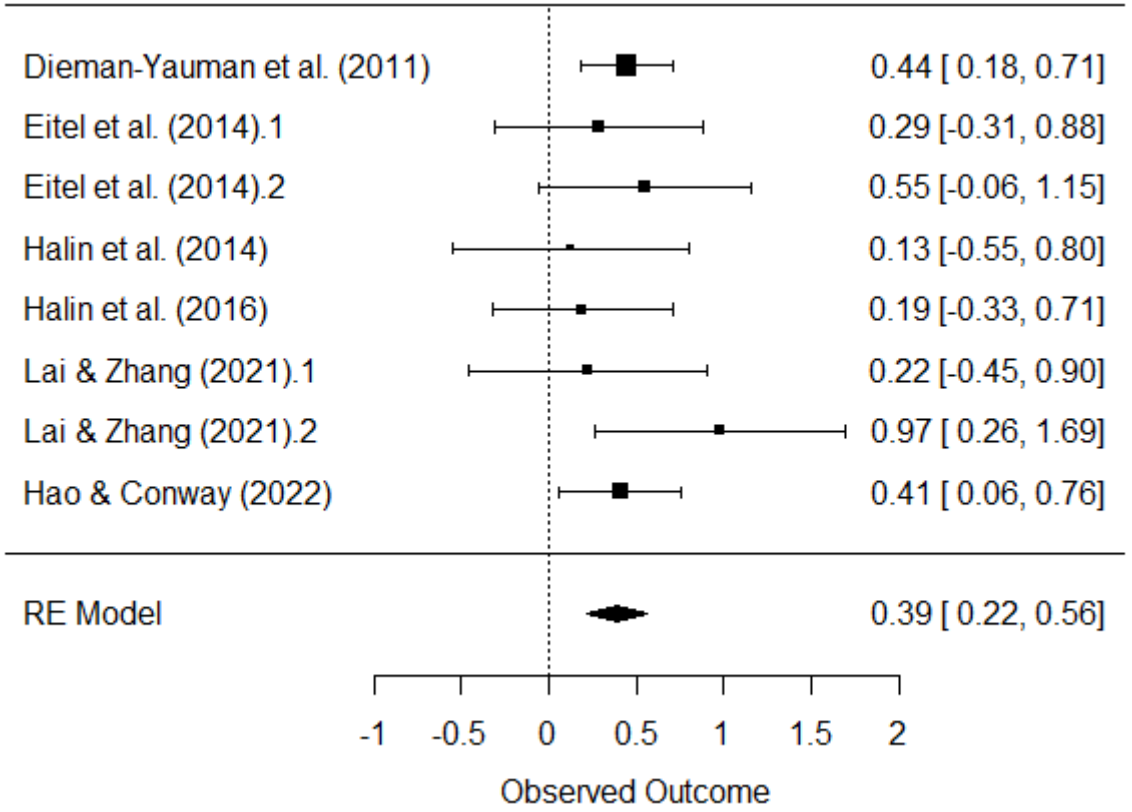
```
library(metafor)
library(clubSandwich)

V = impute_covariance_matrix(vi=df$vi,
                             cluster=df$Study, r=.70)

fit = rma.mv(yi=eff, V=V, data=df,
                         random=~1|Study, slab=Study)

summary(fit)
```

DYSFLUENCY – Main effect

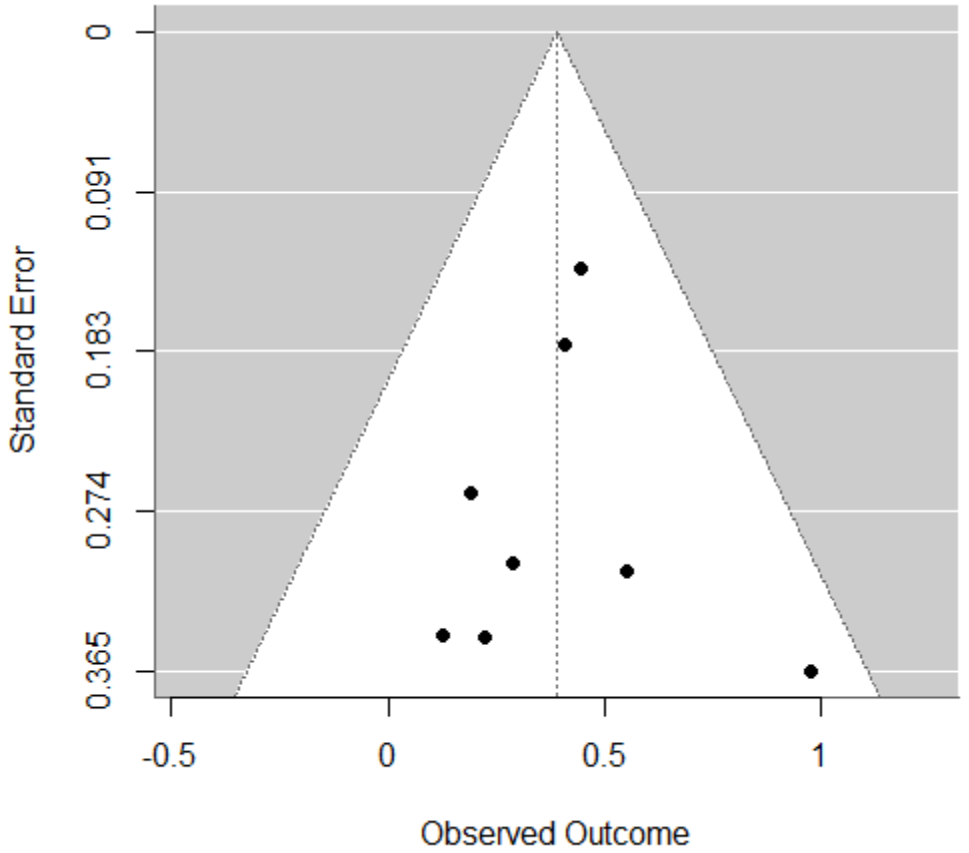
```
Multivariate Meta-Analysis Model (k = 8; method: REML)
  logLik  Deviance       AIC       BIC      AICc
 -1.6818    3.3636    7.3636    7.2554   10.3636

Variance Components:
            estim    sqrt  nlvls  fixed  factor
sigma^2    0.0000  0.0000      6     no   Study

Test for Heterogeneity:
Q(df = 7) = 10.1617, p-val = 0.1796

Model Results:
estimate      se    zval    pval   ci.lb   ci.ub
  0.3901  0.0875  4.4581  <.0001  0.2186  0.5617  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

DYSFLUENCY – Main effect

`forest(fit)`



`funnel(fit)`



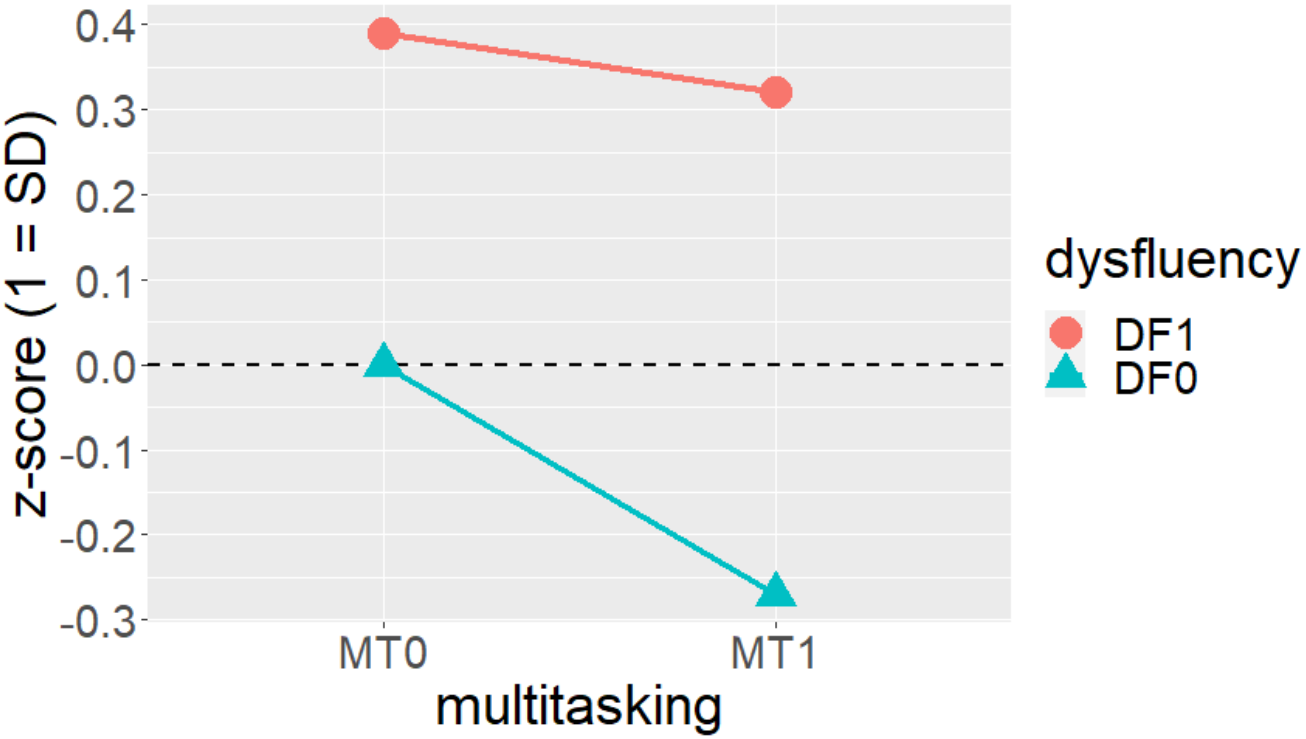
e l'INTERAZIONE?

se l'effetto negativo del *multitasking* è così piccolo (addirittura minore dell'effetto positivo del *dysfluency*) è difficile pensare che possa essere «supercompensato» dal *dysfluency* nel modo inizialmente ipotizzato

si potrebbe ragionare su come rendere il *multitasking* più impattante in modo che l'effetto negativo sia maggiore?

ipotesi semi-plausibile (?) :

```
## set model parameters

intercept = 0
multitaskingEffect = -.28
disfluencyEffect = .39
interactionEffect = .20

sigma = 1
```

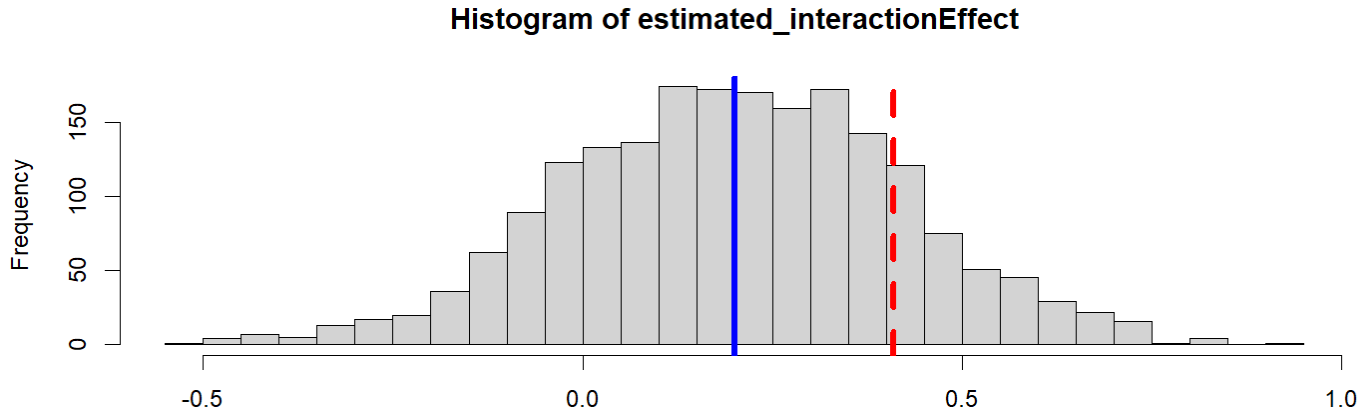


POWER SIMULATION

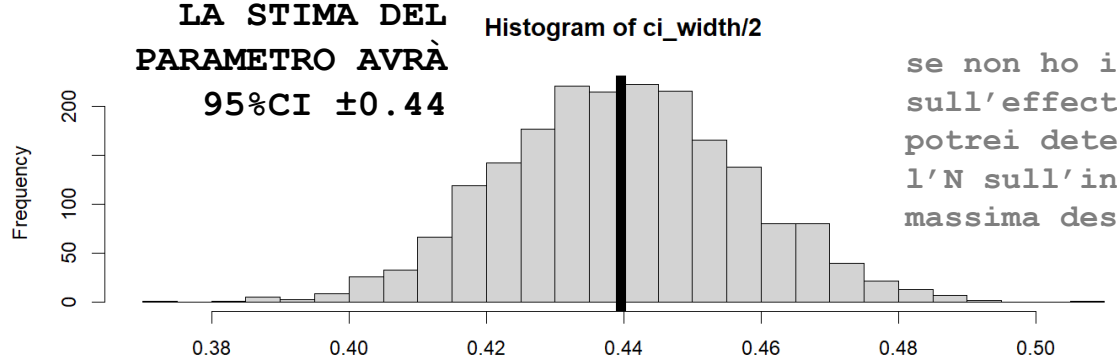
```
N = 320
n = round(N/4)

# actual simulation

niter = 2000
pvalue = rep(NA,niter)
ci_width = rep(NA,niter)
estimated_interactionEffect = rep(NA,niter)

for(i in 1:niter){

  multitasking = c(rep(0,n),rep(0,n),rep(1,n),rep(1,n))
  disfluency   = c(rep(0,n),rep(1,n),rep(0,n),rep(1,n))

  residual = rnorm(N,0,sigma)

  score = intercept +
          multitasking * multitaskingEffect +
          disfluency * disfluencyEffect +
          multitasking * disfluency * interactionEffect +
          residual

  df = data.frame(multitasking, disfluency, score)

  fit1 = lm(score ~ multitasking * disfluency, data=df)

  estimated_interactionEffect[i] =
                        fit1$coefficients["multitasking:disfluency"]

  ci = confint(fit1)
  lower = ci["multitasking:disfluency",1]
  upper = ci["multitasking:disfluency",2]
  ci_width[i] = (upper-lower)

  fit0 = lm(score ~ multitasking + disfluency, data=df)
  pvalue[i] = anova(fit1,fit0)$"Pr(>F)"[2]

  print(round(mean(pvalue < .05, na.rm=T),3))
}
```

```
# POWER con N (tot) = 320
> mean(pvalue < .05, na.rm=T)
[1] 0.134
```

IL MINIMO EFFETTO SIGNIFICATIVO (p<.05) È MAGGIORE DELL'EFFETTO VERO #NONBENE



LA STIMA DEL PARAMETRO AVRÀ 95%CI ±0.44



se non ho ipotesi sull'effect size, potrei determinare l'N sull'incertezza massima desiderata?

COSA RECUPERO CON MISURE RIPETUTE?

Raccogliere misure (risposte) ripetute per soggetto aumenta la potenza? YES, but…

- Il maggiore aumento di potenza si avrebbe trasformando in WITHIN i fattori BETWEEN-SUBJECT e passando ai LMM: concettualmente l'*errore di campionamento* diventa solo quello WITHIN (?, che supponiamo molto minore di quello BETWEEN), con un limite eventualmente dato da *random slope*

- Ripetere le osservazioni mantenendo il disegno BETWEEN x BETWEEN iniziale serve? (ciascun soggetto rimane in una sola cella del disegno) Un po' sì perché riduciamo la *varianza d'errore* dovuta all'*errore di misura*: tanto più 1) quanto più c'è varianza d'errore WITHIN-SUBJECT (*errore di misura?*), 2) quante più osservazioni ripetute raccogliamo. (Oltre un certo limite non serve più)

- 

sarebbe combinare le due cose

ORA RESTIAMO SU BETWEEN x BETWEEN CON MISURE RIPETUTE

Poniamo che la varianza osservata degli *score* sia 1.00 (SD = 1.00, z-score).

Se NON ci fosse alcun *errore di misura*, ripetere lo stesso compito, nelle stesse condizioni, sullo stesso soggetto, darebbe sempre lo stesso *score*. Se invece c'è *errore di misura* (e ce n'è), ogni volta ottengo *score* un po' diversi, ma sperabilmente correlati (chiamo questa correlazione ICC).

Se la misura è buona, la variabilità TRA soggetti è >> di quella ENTRO soggetti, ma comunque entrambe sono < 1.00 (varianza totale). Concettualmente, se posso ridurre la *varianza d'errore* controllando quella ENTRO soggetti via misure ripetute o maggiore attendibilità di misura, è come se aumentassi l'*effect size*.

Avevamo già visto questa cosa a settembre...

Lo z-score è su una metrica in cui l'unità (SD) riflette il totale di variabilità TRA + ENTRO individui

$$\sigma_{TOT} = \sqrt{\sigma^2_{betw} + \sigma^2_{with}} = 1.00$$

Volendo, ad esempio, fissare σ_{TOT} = 1.00 e ICC = 0.60, posso determinare σ_{ID} e σ_{res} risolvendo il sistema di equazioni:

$$\begin{cases} \sqrt{\sigma^2_{betw} + \sigma^2_{with}} = 1.00 \\ \dfrac{\sigma^2_{betw}}{\sigma^2_{betw} + \sigma^2_{with}} = 0.60 \end{cases} \text{ da cui } \begin{cases} \sigma_{with} = 0.632 \\ \sigma_{betw} = 0.775 \end{cases}$$

La variabilità «vera» tra soggetti (σ_{betw}) è inferiore alla variabilità totale dei punteggi

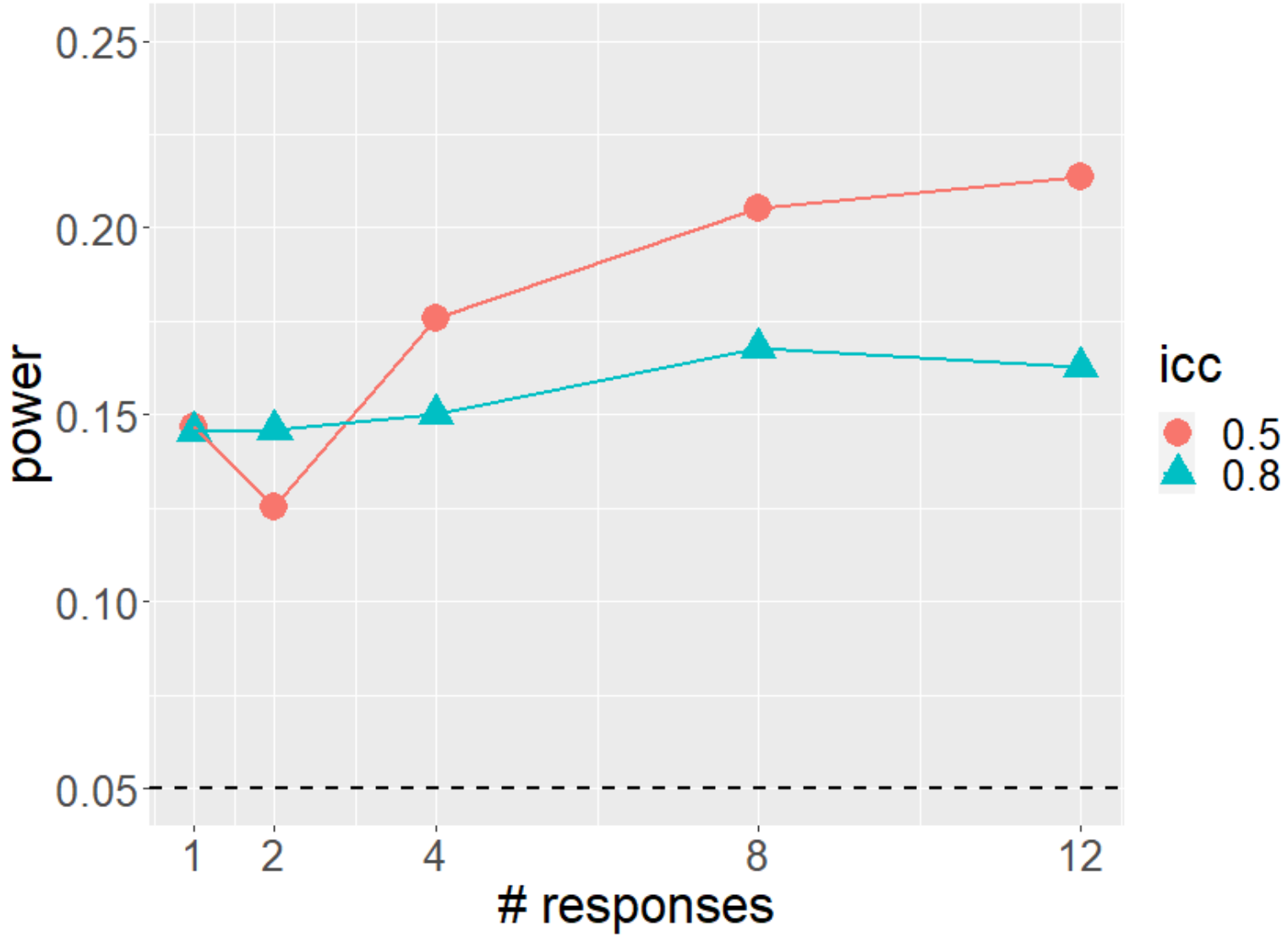
σ_{TOT}	icc	$\sigma_{between}$	σ_{within}
1.00	0.00	0.000	1.000
1.00	0.05	0.224	0.975
1.00	0.10	0.316	0.949
1.00	0.15	0.387	0.922
1.00	0.20	0.447	0.894
1.00	0.25	0.500	0.866
1.00	0.30	0.548	0.837
1.00	0.35	0.592	0.806
1.00	0.40	0.632	0.775
1.00	0.45	0.671	0.742
1.00	0.50	0.707	0.707
1.00	0.55	0.742	0.671
1.00	0.60	0.775	0.632
1.00	0.65	0.806	0.592
1.00	0.70	0.837	0.548
1.00	0.75	0.866	0.500
1.00	0.80	0.894	0.447
1.00	0.85	0.922	0.387
1.00	0.90	0.949	0.316
1.00	0.95	0.975	0.224
1.00	1.00	1.000	0.000

σ_{TOT} fissata a 1 in analogia coi punti z

icc va intesa come la correlazione tra due punteggi forniti dallo stesso soggetto nelle stesse identiche condizioni in momenti diversi (un riferimento è alla *test-retest correlation?*)

set di valori grossomodo plausibili per molte misure psicometriche di prestazione cognitiva

QUINDI QUANTO POWER GUADAGNO? (RESTANDO COL BETWEEN x BETWEEN)



il guadagno di power ottenuto con misure ripetute è più apprezzabile quando l'ICC della misura è minore

la figura è parzialmente fuorviante, perché in realtà ICC maggiore dovrebbe partire dall'inizio con power più alto, ma qui ho fissato sempre la varianza totale a 1 e l'effetto a 0.20 (con ICC maggiore dovrei avere meno varianza totale [perché c'è meno varianza within] oppure *effect size* più grande [se rapportato a una varianza between maggiore])

QUANTO POWER GUADAGNEREI CON TUTTI EFFETTI WITHIN?

```
N = 320
sigmaB = 0.775
sigmaW = 0.632 # icc = 0.60

niter = 1000
pvalue = rep(NA,niter)
estimated_interactionEffect = rep(NA,niter)

for(i in 1:niter){
  rInt = rep(rnorm(N,0,sigmaB),times=4)
  id = rep(1:N,times=4)
  multitasking = c(rep(0,N),rep(0,N),rep(1,N),rep(1,N))
  dysfluency   = c(rep(0,N),rep(1,N),rep(0,N),rep(1,N))
  residual = rnorm(N*4,0,sigmaW)
  score =  intercept +
             multitasking * multitaskingEffect +
             dysfluency * dysfluencyEffect +
             multitasking * dysfluency * interactionEffect +
             rInt + residual
  df = data.frame(id, multitasking, dysfluency, score)
  fit1 = lmer(score ~ multitasking * dysfluency + (1|id), data=df, REML=F)
  estimated_interactionEffect[i] = fit1@beta[4]
  fit0 = lmer(score ~ multitasking + dysfluency + (1|id), data=df, REML=F)
  pvalue[i] = anova(fit1,fit0)$"Pr(>Chisq)"[2]
}
# Power
mean(pvalue < .05, na.rm=T)

# Minimum significant effect
minsign = min(estimated_interactionEffect[estimated_interactionEffect>0 &
                                          pvalue<.05], na.rm=T)
round(minsign,2)

# Distribution of estimated coefficients
hist(estimated_interactionEffect, breaks=20)
abline(v=minsign, col="red",lwd=4,lty=2)
abline(v=interactionEffect, col="blue",lwd=4,lty=1)
```

```
# POWER con N (tot) = 320 e ICC = 0.60
> mean(pvalue < .05, na.rm=T)
[1] 0.797 (prima era 0.134!!!)
```

con ICC maggiore andrebbe ancora meglio

ORA IL MINIMO EFFETTO SIGNIFICATIVO (p<.05) È BEN MINORE DELL'EFFETTO VERO #MOLTOBENE