

Bruno Augusto Soares Gallani

RA: 11202230282

Julia Kohn

RA: 11202130099

Rebeca Oliveira de Souza

RA: 11202230597



Universidade Federal do ABC

Bioestatística – Turma NA2

Professor Rodrigo Pavão

Projeto em grupo - Versão 3

Informações iniciais

- **Título do artigo:** “*Optimizing self-organized study orders: combining refutations and metacognitive prompts improves the use of interleaved practice*”
- **Link do artigo:** <https://www.nature.com/articles/s41539-024-00245-7#Sec9>
- **Dados:** https://osf.io/6vm7z/?view_only=d9caf564f48b4c4f8166b3c2e38c1cf4
- **Dicionário dos dados:** <https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/U9YSER>

Contextualização

O artigo discute sobre a organização de estudos durante a aprendizagem por categorias e suas práticas: a prática intercalada (*interleaved practice*), que consiste na alternância de estudo entre diferentes categorias/conceitos, e a prática por blocos (*blocked practice*), que consiste no estudo de uma categoria/conceito por vez.

A prática de estudo por blocos costuma ser frequentemente escolhida pelos estudantes, em detrimento da prática intercalada, mas pesquisas recentes passaram a informar que a prática intercalada é uma estratégia mais efetiva. Nesse contexto, diversos estudos são realizados a partir de testes com estudantes, testes esses que predefinem uma ordem de estudos. Mas nunca houve um teste que fosse realizado de forma que os próprios estudantes definissem uma ordem de estudos para eles mesmos.

Nesse sentido, foi realizado neste artigo um experimento com 91 estudantes universitários, que consistiu nas seguintes etapas:

- Divisão dos estudantes em grupos controle e teste;
- Exposição dos estudantes do grupo de teste a intervenções: que consistiram na leitura de textos e avisos metacognitivos que destacam a superioridade da prática intercalada de estudos em detrimento à prática por blocos, ou seja, refutações do que comumente acreditam.

Análises realizadas/resultados obtidos

Para analisar os dados obtidos na pesquisa, foram elaboradas três RQs (*good research questions*):

RQ1. Learning strategy beliefs across time and between conditions

Como a crença dos participantes nas diferentes estratégias de aprendizado (em blocos/intercalada) variou com o tempo e entre as condições?

Esta RQ avaliou a influência da intervenção das crenças nas estratégias de aprendizagem (intercalada e por blocos). Para isso, foram aplicados três testes estatísticos:

- **Teste t para amostras pareadas:** para analisar a crença pré existente dos participantes;
- **ANOVA mista:** para observar a variância ao longo do tempo, separando por grupo intervenção e grupo controle;
- **Comparação pareada com Bonferroni.**

Para responder à cada uma das RQs, primeiro faremos um *setup* inicial (importação de bibliotecas necessárias e leitura dos dados).

```
# Importação de bibliotecas
library(afex)
library(emmeans)

# Leitura dos dados
url_csv = paste0(
  "https://docs.google.com/spreadsheets/d/",
  "1Fv6oRXMtoDKcbS19awogukMpTLr5utXLJ3aw8GXzGbA/",
  "export?format=csv&gid=2113233241")

# Ocorreu erro experimental nos dados dos participantes de id 52 e 69,
# e por isso eles possuem valores nulos nas colunas PE_B2 e PE_I2.
dados = read.csv(file = url_csv, sep=",", header=TRUE)
```

Agora, seguindo para a aplicação dos testes:

1.1 - Teste t para amostras pareadas

```
# RQ1 - Crenças pré existentes

# Percepção efetiva *blocked practice*
PE_B1 = dados$PE_B1

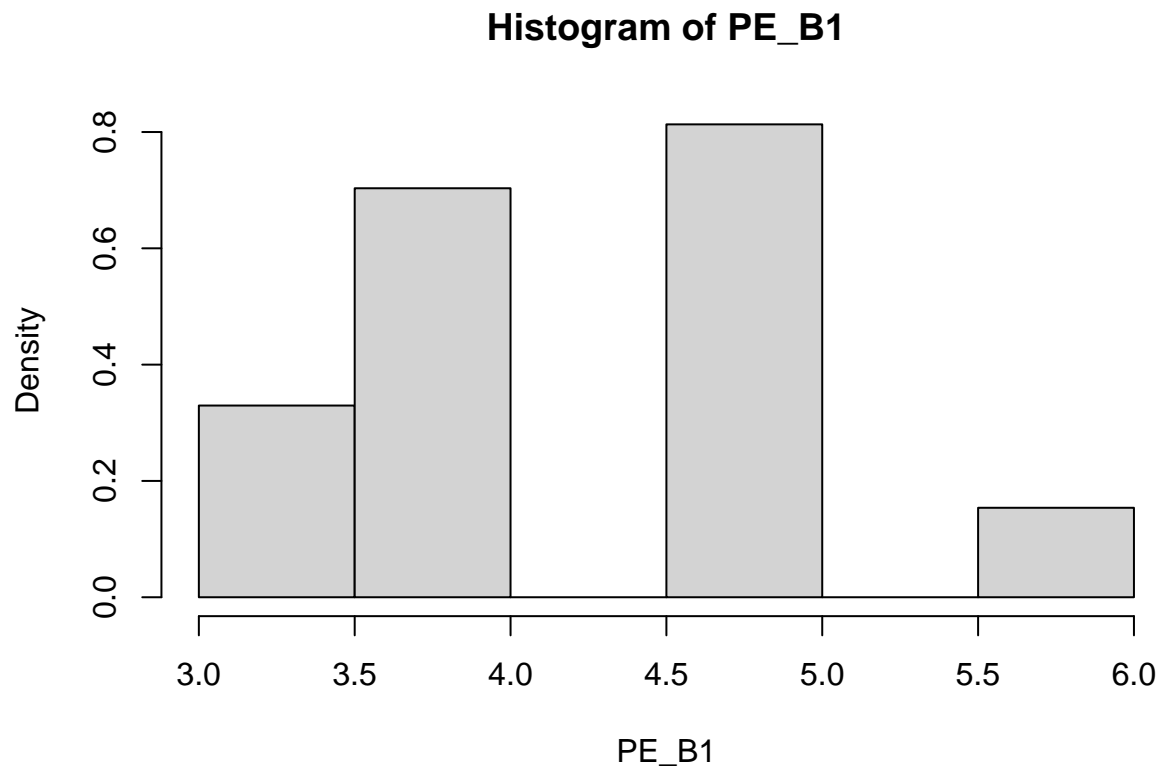
# Média e desvio padrão
c(floor(mean(PE_B1)*100)/100, round(sd(PE_B1), 2))
```

```
## [1] 4.39 0.85
```

```
# A amostra não passou no teste de normalidade
shapiro.test(PE_B1)
```

```
##
## Shapiro-Wilk normality test
##
## data:  PE_B1
## W = 0.86596, p-value = 1.433e-07
```

```
# Observando a distribuição
hist(PE_B1, prob=TRUE)
```



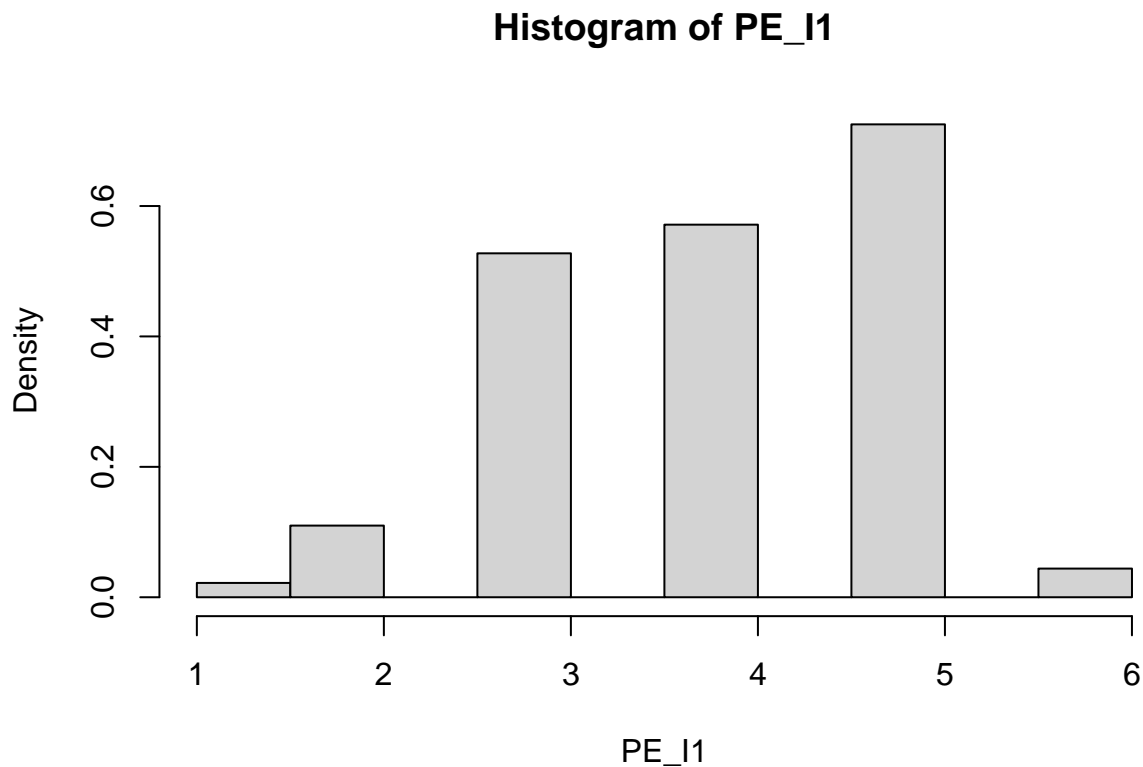
```
# Percepção efetiva *interleaved practice*  
PE_I1 = dados$PE_I1  
  
# Média e desvio padrão  
c(round(mean(PE_I1), 3), round(sd(PE_I1), 2))
```

```
## [1] 4.00 1.02
```

```
# A amostra não passou no teste de normalidade  
shapiro.test(PE_I1)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: PE_I1  
## W = 0.88318, p-value = 6.869e-07
```

```
# Observando a distribuição  
hist(PE_I1, prob=TRUE)
```



```
# teste t pareado
```

```
# H0:  $\mu(PE_{B1}) = \mu(PE_{I1})$ 
```

```
# Ha:  $\mu(PE_{B1}) \neq \mu(PE_{I1})$ 
```

```
t.test(PE_B1, PE_I1, mu=0, paired=TRUE)
```

```
##
## Paired t-test
##
## data: PE_B1 and PE_I1
## t = 2.3619, df = 90, p-value = 0.02034
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.06285266 0.72835613
## sample estimates:
## mean difference
##      0.3956044
```

Pelos resultados do *teste t*, a hipótese nula é rejeitada, devido à obtenção de um **p-valor** (0.020) menor que $\alpha = 0.05$. Mas, considerando que ambas as amostras (PE_{B1} e PE_{I1}) não passaram no teste de normalidade, será que encontraríamos esta mesma conclusão caso aplicássemos a metodologia correta (utilização de um teste não paramétrico)?

```
# Teste de hipótese de duas amostras não paramétrico
```

```
# Como ambas as amostras, PE_B1 e PE_I1, não passaram no teste de normalidade,
```

```
# será utilizado o método não paramétrico de permutação
```

```
# H0: mu(PE_B1) = mu(PE_I1)
# Ha: mu(PE_B1) != mu(PE_I1)

medidas_RQ1 = c(PE_B1, PE_I1)
grupos_RQ1 = c(rep("PE_B1", length(PE_B1)), rep("PE_I1", length(PE_I1)))

# Diferença observada entre as médias
(delta_medias_obs_RQ1 = abs(mean(
  medidas_RQ1[grupos_RQ1=="PE_B1"]) - mean(
  medidas_RQ1[grupos_RQ1=="PE_I1"])))
```

```
## [1] 0.3956044
```

```
# Distribuição das médias regida pela hipótese nula (acaso)
delta_medias_permutacao_RQ1 = vector()
for (p in 1:10000)
{
  grupos_perm_RQ1 = sample(grupos_RQ1, length(grupos_RQ1), replace=FALSE)
  delta_medias_permutacao_RQ1[p] = abs(
    mean(medidas_RQ1[grupos_perm_RQ1=="PE_B1"]) - mean(
    medidas_RQ1[grupos_perm_RQ1=="PE_I1"]))
}

# Cálculo do p-valor para o teste de hipóteses
(pvalor_permutacao_RQ1 = 2*mean(
  delta_medias_permutacao_RQ1 >= delta_medias_obs_RQ1))
```

```
## [1] 0.0114
```

Pelo uso do teste não paramétrico (permutação), obtemos as mesmas conclusões em relação àquelas obtidas pelo uso inadequado do teste `t` (sem validação da premissa de normalidade):

- Rejeição da hipótese nula, devido à obtenção de um **p-valor** menor que $\alpha = 0.05$.

1.2 - ANOVA: variância ao longo do tempo - usando o pacote `afex`

Como a crença nas diferentes técnicas de estudo variou com o tempo/condição?

Setup

```
# Transformação de dados
dados_limpos = na.omit(dados)

dados_ANOVA_RQ1 = data.frame(
  rep(dados_limpos$DM08_01, 3),
  c(rep('pre-intervention', 89),
    rep('post-intervention', 89), rep('after a delay', 89)),
  c(dados_limpos$PE_B1, dados_limpos$PE_B2, dados_limpos$PE_B3),
  c(dados_limpos$PE_I1, dados_limpos$PE_I2, dados_limpos$PE_I3),
  rep(dados_limpos$Condition, 3)
)
colnames(dados_ANOVA_RQ1) <- c("id", "tempo", "percep_blocked",
```

```

                                "percep_interleaved", "condicao")

dados_ANOVA_RQ1_filtro_grupo_intervencao = (
  dados_ANOVA_RQ1[dados_ANOVA_RQ1$condicao == "Full", ])

dados_ANOVA_RQ1_filtro_grupo_controle = (
  dados_ANOVA_RQ1[dados_ANOVA_RQ1$condicao == "Control", ])

# Visualização das 15 primeiras linhas dataframe transformado

head(dados_ANOVA_RQ1, 15)

```

```

##      id      tempo percep_blocked percep_interleaved condicao
## 1    1 pre-intervention          5              3 Control
## 2    2 pre-intervention          6              4 Control
## 3    3 pre-intervention          4              5 Control
## 4    4 pre-intervention          4              3 Control
## 5    5 pre-intervention          4              3 Control
## 6    6 pre-intervention          5              4 Full
## 7    7 pre-intervention          4              4 Control
## 8    8 pre-intervention          5              4 Full
## 9    9 pre-intervention          5              4 Full
## 10  10 pre-intervention          5              3 Full
## 11  11 pre-intervention          5              3 Full
## 12  12 pre-intervention          5              4 Control
## 13  13 pre-intervention          4              5 Control
## 14  14 pre-intervention          4              5 Control
## 15  15 pre-intervention          3              5 Control

```

1.2.1 - Criação dos modelos ANOVA

```

# Modelos ANOVA -- os valores nulos foram desconsiderados

# Variáveis preditoras
# tempo ; condicao

# Variáveis resposta
# percep_blocked; percep_interleaved

# Função para rodar ANOVA mista
rodar_anova <- function(dv, dados, between = TRUE) {
  if (between) {
    aov_ez(id = "id", dv = dv, data = dados, within = "tempo",
            between = "condicao")
  } else {
    aov_ez(id = "id", dv = dv, data = dados, within = "tempo")
  }
}

# Modelos gerais (com fator entre-sujeitos)
modelo_blocked <- rodar_anova("percep_blocked", dados_ANOVA_RQ1)

```

```

## Converting to factor: condicao

```

```
## Contrasts set to contr.sum for the following variables: condicao
```

```
modelo_interleaved <- rodar_anova("percep_interleaved", dados_ANOVA_RQ1)
```

```
## Converting to factor: condicao
## Contrasts set to contr.sum for the following variables: condicao
```

```
# Modelos separados por grupo (sem fator entre-sujeitos)
modelo_blocked_interv <- rodar_anova("percep_blocked",
                                     dados_ANOVA_RQ1_filtro_grupo_intervencao,
                                     between = FALSE)

modelo_blocked_control <- rodar_anova("percep_blocked",
                                     dados_ANOVA_RQ1_filtro_grupo_controle,
                                     between = FALSE)

modelo_interleaved_interv <- rodar_anova(
  "percep_interleaved",
  dados_ANOVA_RQ1_filtro_grupo_intervencao,
  between = FALSE)

modelo_interleaved_control <- rodar_anova(
  "percep_interleaved",
  dados_ANOVA_RQ1_filtro_grupo_controle,
  between = FALSE)
```

1.2.2 - Outputs dos modelos

- Prática de estudos em bloco

```
# ANOVA mista - Análise da percepção dos participantes referente à prática
# de estudos em bloco
```

```
# Tempo: possui efeitosignificativo (p-valor < 0.05)
# Condição: não possui efeito significativo (p-valor > 0.05)
# Interação tempo x condição: possui efeito significativo (p-valor < 0.05)
```

```
modelo_blocked
```

```
## Anova Table (Type 3 tests)
##
## Response: percep_blocked
##          Effect      df  MSE      F ges p.value
## 1      condicao      1, 87 1.53    2.23 .015   .139
## 2      tempo 1.76, 153.11 0.64 22.65 *** .100  <.001
## 3 condicao:tempo 1.76, 153.11 0.64  8.20 *** .039  <.001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```



```
# ANOVA mista - Análise da percepção dos participantes referente à prática
# de estudos em bloco, olhando apenas para o grupo de intervenção

# H0: a influência do tempo, nos participantes do grupo de intervenção,
# na percepção de eficácia da prática de estudos em bloco é dada ao acaso

# Ha: nos participantes do grupo de intervenção, o tempo teve influência
# na percepção de eficácia da prática de estudos em bloco

# p-valor < 0.05: rejeição da hipótese nula
modelo_blocked_interv
```

```
## Anova Table (Type 3 tests)
##
## Response: percep_blocked
##      Effect      df  MSE      F ges p.value
## 1 tempo 1.89, 81.27 0.60 28.24 *** .229    <.001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

```
# ANOVA mista - Análise da percepção dos participantes referente à prática
# de estudos em bloco, olhando apenas para o grupo controle

# H0: a influência do tempo, nos participantes do grupo controle,
# na percepção de eficácia da prática de estudos em bloco é dada ao acaso

# Ha: nos participantes do grupo controle, o tempo teve influência
# na percepção de eficácia da prática de estudos em bloco

# p-valor > 0.05: aceitação da hipótese nula
modelo_blocked_control
```

```
## Anova Table (Type 3 tests)
##
## Response: percep_blocked
##      Effect      df  MSE      F ges p.value
## 1 tempo 1.54, 67.67 0.73 2.00 .018    .153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

- Prática de estudos intercalada

```
# ANOVA mista - Análise da percepção dos participantes referente à prática
# de estudos intercalada

# Tempo: possui efeito significativo (p-valor < 0.05)
# Condição: possui efeito significativo (p-valor < 0.05)
# Interação tempo x condição: possui efeito significativo (p-valor < 0.05)
```

modelo_interleaved

```
## Anova Table (Type 3 tests)
##
## Response: percep_interleaved
##          Effect          df  MSE          F ges p.value
## 1      condicao           1, 87 1.60   6.29 * .038      .014
## 2          tempo 1.62, 140.70 0.84   6.44 ** .033      .004
## 3 condicao:tempo 1.62, 140.70 0.84   4.32 * .022      .022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

*# ANOVA mista - Análise da percepção dos participantes referente à prática
de estudos intercalada, olhando apenas para o grupo de intervenção*

*# H0: a influência do tempo, nos participantes do grupo de intervenção,
na percepção de eficácia da prática de estudos intercalada é dada ao acaso*

*# Ha: nos participantes do grupo de intervenção, o tempo teve influência
na percepção de eficácia da prática de estudos intercalada*

modelo_interleaved_interv

```
## Anova Table (Type 3 tests)
##
## Response: percep_interleaved
##          Effect          df  MSE          F ges p.value
## 1      tempo 1.57, 67.44 0.74 11.62 *** .115    <.001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

*# ANOVA mista - Análise da percepção dos participantes referente à prática
de estudos intercalada, olhando apenas para o grupo controle*

*# H0: a influência do tempo, nos participantes do grupo controle,
na percepção de eficácia da prática de estudos intercalada é dada ao acaso*

*# Ha: nos participantes do grupo controle, o tempo teve influência
na percepção de eficácia da prática intercalada*

modelo_interleaved_control

```
## Anova Table (Type 3 tests)
##
## Response: percep_interleaved
##   Effect      df  MSE    F ges p.value
## 1  tempo 1.64, 72.17 0.95 0.69 .007    .477
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

1.3 - Comparações pareadas com *Bonferroni*

- Prática de estudos em bloco

```
# Pós-intervenção
dados_post = subset(dados_ANOVA_RQ1, tempo == "post-intervention")

# Valores de média e desvio padrão para o grupo controle
media_percepcao_eficacia_pratica_bloco_grupo_controle_post = mean(na.omit(
  dados_post$percep_blocked[dados_post$condicao == "Control"]))

desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_controle_post = sd(
  na.omit(
    dados_post$percep_blocked[dados_post$condicao == "Control"]))

c(round(media_percepcao_eficacia_pratica_bloco_grupo_controle_post, 2),
  round(desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_controle_post, 2))
```

```
## [1] 3.91 1.10
```

```
# Valores de média e desvio padrão para o grupo de intervenção
media_percepcao_eficacia_pratica_bloco_grupo_intervencao_post = mean(na.omit(
  dados_post$percep_blocked[dados_post$condicao == "Full"]))

desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_intervencao_post = sd(
  na.omit(
    dados_post$percep_blocked[dados_post$condicao == "Full"]))

c(round(media_percepcao_eficacia_pratica_bloco_grupo_intervencao_post, 2),
  round(
    desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_intervencao_post, 2))
```

```
## [1] 3.43 0.95
```

Após a intervenção, a percepção da eficácia da prática de estudos em blocos,
pelos participantes, foi diferente entre os grupos controle e intervenção,
sendo maior no grupo controle.

```
(t_post_blocked = t.test(percep_blocked ~ condicao,
  data = dados_post, var.equal = TRUE) )
```

```
##
## Two Sample t-test
##
## data: percep_blocked by condicao
## t = 2.1932, df = 87, p-value = 0.03096
## alternative hypothesis: true difference in means between group Control and group Full is
## ↪ not equal to 0
## 95 percent confidence interval:
##  0.04491944 0.91366642
## sample estimates:
## mean in group Control      mean in group Full
##           3.911111           3.431818
```

```
# Após o delay
dados_delay = subset(dados_ANOVA_RQ1, tempo == "after a delay")

# Valores de média e desvio padrão para o grupo controle
media_percepcao_eficacia_pratica_bloco_grupo_controle_delay = mean(na.omit(
  dados_delay$percep_blocked[dados_post$condicao == "Control"]))

desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_controle_delay = sd(
  na.omit(
    dados_delay$percep_blocked[dados_post$condicao == "Control"]))
)

c(round(media_percepcao_eficacia_pratica_bloco_grupo_controle_delay, 2),
  round(desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_controle_delay, 2))
```

```
## [1] 4.02 0.94
```

```
# Valores de média e desvio padrão para o grupo de intervenção
media_percepcao_eficacia_pratica_bloco_grupo_intervencao_delay = mean(na.omit(
  dados_delay$percep_blocked[dados_post$condicao == "Full"]))

desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_intervencao_delay = sd(
  na.omit(
    dados_delay$percep_blocked[dados_post$condicao == "Full"]))

c(round(media_percepcao_eficacia_pratica_bloco_grupo_intervencao_delay, 2),
  round(
    desvio_padrao_percepcao_eficacia_pratica_bloco_grupo_intervencao_delay, 3))
```

```
## [1] 3.520 0.952
```

```
# Após um período pós-intervenção, a percepção da eficácia da prática
# de estudos em blocos, pelos participantes, foi diferente entre os grupos
# controle e intervenção, sendo novamente maior no grupo controle.

(t_delay_blocked = t.test(percep_blocked ~ condicao,
  data = dados_delay, var.equal = TRUE) )
```

```
##
## Two Sample t-test
##
## data: percep_blocked by condicao
## t = 2.4889, df = 87, p-value = 0.01472
## alternative hypothesis: true difference in means between group Control and group Full is
## ↪ not equal to 0
## 95 percent confidence interval:
##  0.1006026 0.8983873
## sample estimates:
## mean in group Control      mean in group Full
##           4.022222           3.522727
```

- Prática de estudos intercalada

Pós intervenção

Valores de média e desvio padrão para o grupo controle

```
media_percepcao_eficacia_pratica_intercal_grupo_controle_post = mean(na.omit(
  dados_post$percep_interleaved[dados_post$condicao == "Control"]))
```

```
desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_controle_post = sd(
  na.omit(dados_post$percep_interleaved[dados_post$condicao == "Control"]))
```

```
c(round(media_percepcao_eficacia_pratica_intercal_grupo_controle_post, 2),
  round(desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_controle_post, 2))
```

```
## [1] 4.04 1.04
```

Valores de média e desvio padrão para o grupo de intervenção

```
media_percepcao_eficacia_pratica_intercal_grupo_intervencao_post = mean(
  na.omit(
    dados_post$percep_interleaved[dados_post$condicao == "Full"]))
```

```
desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_intervencao_post = sd(
  na.omit(dados_post$percep_interleaved[dados_post$condicao == "Full"]))
```

```
c(round(media_percepcao_eficacia_pratica_intercal_grupo_intervencao_post, 2),
  round(
    desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_intervencao_post,
    2))
```

```
## [1] 4.80 0.79
```

Após a intervenção, a percepção da eficácia da prática de estudos intercalada, pelos participantes, foi diferente entre os grupos controle e intervenção, sendo maior no grupo de intervenção.

```
(t_post_interleaved = t.test(percep_interleaved ~ condicao,
  data = dados_post, var.equal = TRUE) )
```

```
##
## Two Sample t-test
##
## data: percep_interleaved by condicao
## t = -3.8134, df = 87, p-value = 0.0002556
## alternative hypothesis: true difference in means between group Control and group Full is
## ↪ not equal to 0
## 95 percent confidence interval:
## -1.1424521 -0.3595681
## sample estimates:
## mean in group Control      mean in group Full
##           4.044444           4.795455
```

Depois de um delay

Valores de média e desvio padrão para o grupo controle

```
media_percepcao_eficacia_pratica_intercal_grupo_controle_delay = mean(na.omit(
  dados_delay$percep_interleaved[dados_delay$condicao == "Control"]))

desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_controle_delay = sd(
  na.omit(dados_delay$percep_interleaved[dados_delay$condicao == "Control"]))

c(round(media_percepcao_eficacia_pratica_intercal_grupo_controle_delay, 2),
  round(
    desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_controle_delay, 2))
```

```
## [1] 4.22 1.13
```

Valores de média e desvio padrão para o grupo de intervenção

```
media_percepcao_eficacia_pratica_intercal_grupo_intervencao_delay = mean(
  na.omit(
    dados_delay$percep_interleaved[dados_delay$condicao == "Full"]))

desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_intervencao_delay = sd(
  na.omit(
    dados_delay$percep_interleaved[dados_delay$condicao == "Full"]))

c(round(media_percepcao_eficacia_pratica_intercal_grupo_intervencao_delay, 2),
  round(
    desvio_padrao_percepcao_eficacia_pratica_intercal_grupo_intervencao_delay,
    2))
```

```
## [1] 4.61 0.95
```

*# Após um período pós-intervenção, a percepção da eficácia da prática de
estudos intercalada, pelos participantes, foi diferente entre os grupos
controle e intervenção, sendo novamente maior no grupo de intervenção.*

```
(t_delay_interleaved = t.test(percep_interleaved ~ condicao,
  data = dados_delay, var.equal = TRUE) )
```

```
##
## Two Sample t-test
##
## data: percep_interleaved by condicao
## t = -1.7741, df = 87, p-value = 0.07954
## alternative hypothesis: true difference in means between group Control and group Full is
## ↪ not equal to 0
## 95 percent confidence interval:
## -0.82992581 0.04709752
## sample estimates:
## mean in group Control    mean in group Full
##           4.222222           4.613636
```

RQ2. The use of interleaved practice across time and between conditions

Como o uso da prática de estudos intercalada variou com o tempo e entre os diferentes grupos (controle e experimental)?

Esta RQ avaliou a influência das variáveis de **tempo** e **condição** no uso da prática de estudos intercalada. Para isso, foram aplicados três testes estatísticos:

- **Teste t para uma amostra:** para analisar se a preferência da prática de estudos em bloco pelos estudantes, em detrimento à prática intercalada, é dada ao acaso;
- **ANOVA mista:** para observar a variância ao longo do tempo do uso da prática de estudo intercalada, separando por grupo intervenção e grupo controle;
- **Comparações pareadas com Bonferroni**

Seguindo para a aplicação dos testes:

2.1 - Teste t para uma amostra

```
# Primeiramente, antes de aplicar o teste t, verificaremos
# a diferença entre as frequências de uso entre as
# duas práticas de estudo (por blocos e intercalada)

# Proporção de trocas
freq_interleaved_trials = mean(dados$`sw.1`)

freq_blocked_trials = 1 - freq_interleaved_trials

c(round(freq_blocked_trials, 2),
  round(freq_interleaved_trials, 2))
```

```
## [1] 0.57 0.43
```

Pelos dados obtidos, observamos que a frequência de uso das práticas de estudo em bloco é maior que a frequência de uso das práticas de estudo intercalado.

Mas esta diferença nas frequências seria dada ao acaso? Isto é, o uso da prática de estudos em bloco é intencional?

```

# Aplicação do teste t
# Dados de preferências pela prática de estudos em blocos pelos estudantes
# vai de 0 a 1
blocked_trial_preference_pre_intervention = 1 - dados$`sw.1`

# Média e desvio padrão
c(mean(blocked_trial_preference_pre_intervention),
sd(blocked_trial_preference_pre_intervention))

```

```
## [1] 0.5682889 0.2993402
```

```

# Verificar se atende a premissa
# Não passou no teste de normalidade
shapiro.test(blocked_trial_preference_pre_intervention)

```

```

##
##  Shapiro-Wilk normality test
##
## data:  blocked_trial_preference_pre_intervention
## W = 0.84428, p-value = 2.341e-08

```

```

# Teste t para uma amostra
# H0:  $\mu = 0.16$  (a média observada é igual a 0,16)
# Ha:  $\mu \neq 0.16$  (a média observada é diferente de 0,16)

# O valor 0.16 refere-se ao valor esperado para a escolha
# de uma categoria aleatória.
# Como o estudo foi realizado com 6 diferentes categorias,
# então  $E(X) = 1/6$  (aprox. 0.16).

muH0 = 0.16
t.test(blocked_trial_preference_pre_intervention, mu = muH0)

```

```

##
##  One Sample t-test
##
## data:  blocked_trial_preference_pre_intervention
## t = 13.011, df = 90, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.16
## 95 percent confidence interval:
##  0.5059482 0.6306295
## sample estimates:
## mean of x
## 0.5682889

```

Rejeitamos a hipótese nula (H_0) de que a média do uso de práticas de estudo em blocos é igual a 0.16.

A média observada (0.57) é significativamente diferente de 0.16, com um p-valor muito baixo ($< 2.2 \times 10^{-16}$), indicando que o uso de práticas de estudo em blocos na pré-intervenção não pode ser explicado apenas pelo acaso. Mas, considerando que a amostra `blocked_trial_preference_pre_intervention` não passou no teste de normalidade, será que encontraríamos esta mesma conclusão caso aplicássemos a metodologia correta (utilização de um teste não paramétrico)?


```

# Teste de hipótese não paramétrico para uma amostra
# Como a amostra "blocked_trial_preference_pre_intervention" não passou
# no teste de normalidade, será utilizado o método não paramétrico de bootstrap

# H0:  $\mu = 0.16$  (a média observada é igual a 0,16)
# Ha:  $\mu \neq 0.16$  (a média observada é diferente de 0,16)

# O valor 0.16 refere-se ao valor esperado para a escolha
# de uma categoria aleatória.
# Como o estudo foi realizado com 6 diferentes categorias,
# então  $E(X) = 1/6$  (aprox. 0.16).

# Amostra --> dados de preferências pela prática
# de estudos em blocos pelos estudantes

amostra_RQ2 = blocked_trial_preference_pre_intervention
media_BS_RQ2 = vector()
for(i in 1:100000)
{
  amostra_BS_RQ2 = sample(amostra_RQ2, length(amostra_RQ2), replace=TRUE)
  media_BS_RQ2[i] = mean(amostra_BS_RQ2)
}

# Pela análise do intervalo de confiança, conseguimos observar que o valor
# esperado pela hipótese nula (0.16) não faz parte do intervalo.
# Portanto, rejeitaríamos a hipótese nula
cat("Intervalo de confiança 95% da média amostral:",
    '\n\n', quantile(media_BS_RQ2, c(0.025, 0.975)))

```

```

## Intervalo de confiança 95% da média amostral:
##
## 0.5061224 0.6285714

```

Pelo uso do teste não paramétrico (*bootstrap*), obtemos as mesmas conclusões em relação àquelas obtidas pelo uso inadequado do teste *t* (sem validação da premissa de normalidade):

- Rejeição da hipótese nula, uma vez que o intervalo de confiança 95% da média amostral [0.506, 0.628] não compreende o valor esperado pela hipótese nula (0.16).

2.2 - ANOVA mista

Antes de utilizar o modelo de ANOVA, é necessário preparar os dados:

```

# Dados transformados para a ANOVA da RQ2
dados_ANOVA_RQ2 = data.frame(
  rep(dados$DM08_01, 3),
  c(rep('pre-intervention', 91), rep('post-intervention', 91),
    rep('after a delay', 91)),
  c(dados$nsw_1, dados$nsw_2, dados$nsw_3),
  c(dados$`sw.1`, dados$`sw.2`, dados$`sw.3`),
  c(dados$PE_B1, dados$PE_B2, dados$PE_B3),
  c(dados$PE_I1, dados$PE_I2, dados$PE_I3),
  rep(dados$Condition, 3)
)

```

```
colnames(dados_ANOVA_RQ2) = c("id", "tempo", "numero_trocas",
"proporcao_trocas", "percep_blocked", "percep_interleaved",
"condicao")
```

Agora, conseguimos criar os modelos ANOVA:

```
# Usando a função aov_ez para realizar a ANOVA mista 2x3

# A proporção de trocas refere-se à proporção de
# uso da prática de estudos intercalada

# Função do usuário
modelo_RQ2 = rodar_anova(dv="proporcao_trocas", dados_ANOVA_RQ2)

# Função da biblioteca afex
modelo_RQ2 = aov_ez(id = "id",
                    dv = "proporcao_trocas",
                    within = "tempo",
                    between = "condicao",
                    data = dados_ANOVA_RQ2)
```

```
## Converting to factor: condicao
```

```
## Contrasts set to contr.sum for the following variables: condicao
```

```
# Resumo do modelo
modelo_RQ2
```

```
## Anova Table (Type 3 tests)
##
## Response: proporcao_trocas
##          Effect      df  MSE      F ges p.value
## 1      condicao      1, 89 0.19   7.48 ** .049   .008
## 2      tempo 1.50, 133.58 0.08 30.74 *** .119  <.001
## 3 condicao:tempo 1.50, 133.58 0.08  9.00 *** .038  <.001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

Pelo *output* do modelo e considerando um valor para o **nível de significância** ($\alpha = 5\% = 0.05$), foram obtidos os seguintes resultados:

- A variável de **tempo** exerceu um efeito significativo no uso da prática de estudos intercalada ($p < 0.001$):
 - $F(1.50, 133.58) = 30.74$
 - $p < \alpha \rightarrow$ rejeição da hipótese nula
- A variável de **condição** exerceu um efeito significativo no uso da prática de estudos intercalada ($p = 0.008 < \alpha$):

- $F(1, 89) = 7.48$
 - $p < \alpha \rightarrow$ rejeição da hipótese nula
- A interação das variáveis de **condição** e **tempo** (condição*tempo) exerceu um efeito significativo no uso da prática de estudos intercalada ($p < 0.001$):
 - $F(1.50, 133.58) = 9.00$
 - $p < \alpha \rightarrow$ rejeição da hipótese nula

2.3 - Comparações pareadas com *Bonferroni*

1. O uso da prática intercalada no grupo de intervenção (*full*) após a intervenção foi diferente do grupo controle?
2. Essa diferença também é encontrada após o *delay*?

```
# Converter as variáveis para fatores
dados_ANOVA_RQ2$tempo = factor(dados_ANOVA_RQ2$tempo,
                                levels = c("pre-intervention",
                                             "post-intervention",
                                             "after a delay"))

dados_ANOVA_RQ2$condicao = as.factor(dados_ANOVA_RQ2$condicao)

# Para o momento pós-intervenção:
dados_post = subset(dados_ANOVA_RQ2, tempo == "post-intervention")
resultado_post = pairwise.t.test(dados_post$proporcao_trocas,
                                  dados_post$condicao,
                                  p.adjust.method = "bonferroni")

print(resultado_post)
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: dados_post$proporcao_trocas and dados_post$condicao
##
## Control
## Full 0.0037
##
## P value adjustment method: bonferroni
```

*#Pvalor 0.0037, bem menor que 0.05, então podemos dizer que
há diferença não explicada pelo acaso*

```
# Para o momento após um atraso:
dados_delay = subset(dados_ANOVA_RQ2, tempo == "after a delay")
resultado_delay = pairwise.t.test(dados_delay$proporcao_trocas,
                                   dados_delay$condicao,
                                   p.adjust.method = "bonferroni")

print(resultado_delay)
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: dados_delay$proporcao_trocas and dados_delay$condicao
##
## Control
## Full 3e-04
##
## P value adjustment method: bonferroni
```

*# Pvalor de 0.0003, menor que 0.05, então podemos dizer que
há diferença não explicada pelo acaso*

RQ3. The influence of interleaved practice on classification accuracy

A prática de estudos intercalada possui algum efeito sobre a acurácia de classificação?

- Verificando se os diferentes grupos (controle e experimental) possuem algum efeito sobre a performance de classificação, nas tarefas controladas pelo experimentador (*experimented-controlled learning tasks*).

```
# Converter a variável de condição para fator
dados$Condition = as.factor(dados$Condition)

# Teste t comparando a acurácia das tarefas controladas pelo
# experimentador entre os diferentes grupos (controle e experimental)

teste_control = t.test(test.b + test.i ~ Condition,
                        data = dados, var.equal = TRUE)

teste_control
```

```
##
## Two Sample t-test
##
## data: test.b + test.i by Condition
## t = -0.67901, df = 89, p-value = 0.4989
## alternative hypothesis: true difference in means between group Control and group Full is
## ↪ not equal to 0
## 95 percent confidence interval:
## -2.577687 1.264643
## sample estimates:
## mean in group Control mean in group Full
## 9.80000 10.45652
```

*# Os grupos não possuem efeitos sobre a acurácia das tarefas de aprendizagem
controladas pelo experimentador.*

```
# Teste de hipótese de duas amostras não paramétrico
# Como não foi feita a validação de premissas de normalidade em nenhuma
# das amostras, faremos o uso do teste não paramétrico de permutação

# H0: mu(somaTestes condicao controle) = mu(somaTestes condicao experimental)
```

```

# Ha: mu(somaTestes condicao controle) != mu(somaTestes condicao experimental)

# Dados transformados para o teste t da RQ3 -- permutação
dados_teste_t_perm_RQ3 = data.frame(
  id = dados$DM08_01,
  score_teste_experimental_controlado = dados$test.i + dados$test.b,
  condicao = dados$Condition
)

amostra_teste_t_RQ3_1_controle =
  dados_teste_t_perm_RQ3$score_teste_experimental_controlado[
    dados_teste_t_perm_RQ3$condicao == "Control"
  ]

amostra_teste_t_RQ3_1_experimental =
  dados_teste_t_perm_RQ3$score_teste_experimental_controlado[
    dados_teste_t_perm_RQ3$condicao == "Full"
  ]

medidas_RQ3_1 = c(
  amostra_teste_t_RQ3_1_controle, amostra_teste_t_RQ3_1_experimental)

grupos_RQ3_1 = c(rep("Controle", length(amostra_teste_t_RQ3_1_controle)),
  rep("Experimental", length(amostra_teste_t_RQ3_1_experimental)))

# Diferença observada entre as médias
(delta_medias_obs_RQ3_1 = abs(mean(
  medidas_RQ3_1[grupos_RQ3_1=="Controle"]) - mean(
  medidas_RQ3_1[grupos_RQ3_1=="Experimental"])))

```

```
## [1] 0.6565217
```

```

# Distribuição das médias regida pela hipótese nula (acaso)
delta_medias_permutacao_RQ3_1 = vector()
for (p in 1:10000)
{
  grupos_perm_RQ3_1 = sample(grupos_RQ3_1, length(grupos_RQ3_1), replace=FALSE)
  delta_medias_permutacao_RQ3_1[p] = abs(
    mean(medidas_RQ3_1[grupos_perm_RQ3_1=="Controle"]) - mean(
    medidas_RQ3_1[grupos_perm_RQ3_1=="Experimental"]))
}

# Cálculo do p-valor para o teste de hipóteses
(pvalor_permutacao_RQ3_1 = min(
  2*mean(delta_medias_permutacao_RQ3_1 >= delta_medias_obs_RQ3_1), 1))

```

```
## [1] 1
```

```

# Pelo uso do teste não paramétrico (permutação),
# obtemos as mesmas conclusões em relação àquelas obtidas
# pelo uso inadequado do `teste t` (sem validação da premissa de normalidade):

```

```
# Aceitação da hipótese nula, devido à obtenção de um p-valor (~ 1) maior que  
# alpha = 0.05.
```

- Nas tarefas controladas pelo experimentador, houve diferença na acurácia de classificação entre as diferentes estratégias de aprendizagem?

```
# Teste t pareado comparando a acurácia para prática intercalada (test.i)  
# versus bloqueada (test.b)  
teste_pareado = t.test(dados$test.i, dados$test.b, paired = TRUE)  
print(teste_pareado)
```

```
##  
## Paired t-test  
##  
## data: dados$test.i and dados$test.b  
## t = 10.005, df = 90, p-value = 2.79e-16  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## 2.395484 3.582538  
## sample estimates:  
## mean difference  
## 2.989011
```

```
media_interleaved = mean(dados$test.i, na.rm = TRUE)  
dp_interleaved = sd(dados$test.i, na.rm = TRUE)  
media_blocked = mean(dados$test.b, na.rm = TRUE)  
dp_blocked = sd(dados$test.b, na.rm = TRUE)  
  
cat("Prática intercalada: M =", round(media_interleaved,2), "DP =",  
    round(dp_interleaved,2), "\n")
```

```
## Prática intercalada: M = 6.56 DP = 2.97
```

```
cat("Prática bloqueada: M =", round(media_blocked,2), "DP =",  
    round(dp_blocked,2), "\n")
```

```
## Prática bloqueada: M = 3.57 DP = 2.4
```

```
# Como a diferença média encontrada foi 2.989 e o p-valor foi muito baixo,  
# podemos afirmar que essa diferença não se deve ao acaso.
```

```
# Mas será que as amostras passaram no teste de normalidade?
```

```
# Não passou no teste de normalidade  
shapiro.test(dados$test.b)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: dados$test.b  
## W = 0.9492, p-value = 0.00139
```

```
# Não passou no teste de normalidade
shapiro.test(dados$test.i)
```

```
##
## Shapiro-Wilk normality test
##
## data:  dados$test.i
## W = 0.96457, p-value = 0.01406
```

```
# Teste de hipótese de duas amostras não paramétrico
# Como as amostras não passaram na premissa de normalidade,
# utilizaremos o método não paramétrico de permutação
```

```
# H0:  $\mu(\text{Prática intercalada}) = \mu(\text{Prática bloqueada})$ 
# Ha:  $\mu(\text{Prática bloqueada}) \neq \mu(\text{Prática intercalada})$ 
```

```
n = length(dados$test.i)
medidas_RQ3 = c(dados$test.i, dados$test.b)
grupos_RQ3 = c(rep("Prática intercalada", n), rep("Prática bloqueada", n))
```

```
# Diferença observada entre as médias
(delta_medias_obs_RQ3 = abs(mean(
  medidas_RQ3[grupos_RQ3=="Prática intercalada"]) - mean(
  medidas_RQ3[grupos_RQ3=="Prática bloqueada"])))
```

```
## [1] 2.989011
```

```
# Distribuição das médias regida pela hipótese nula (acaso)
delta_medias_permutacao_RQ3 = vector()
for (p in 1:10000)
{
  grupos_perm_RQ3 = sample(grupos_RQ3, length(grupos_RQ3), replace=FALSE)
  delta_medias_permutacao_RQ3[p] = abs(
    mean(medidas_RQ3[grupos_perm_RQ3=="Prática intercalada"]) - mean(
      medidas_RQ3[grupos_perm_RQ3=="Prática bloqueada"]))
}
```

```
# Cálculo do p-valor para o teste de hipóteses
(pvalor_permutacao_RQ3 = 2*mean(
  delta_medias_permutacao_RQ3 >= delta_medias_obs_RQ3))
```

```
## [1] 0
```

```
# Pelo uso do teste não paramétrico (permutação), obtemos as mesmas
# conclusões em relação àquelas obtidas pelo uso inadequado do teste t
# (sem validação da premissa de normalidade):
```

```
# Rejeição da hipótese nula, devido à obtenção de um p-valor menor que
#  $\alpha = 0.05$ 
```

- Nos casos em que os estudantes definiram suas próprias ordens de estudo, quais variáveis influenciaram a performance de classificação?

Uso de um modelo de ANOVA mista para verificar se as condições de controle e intervenção diferiam em termos de desempenho de classificação

Antes de utilizar o modelo ANOVA, é necessária a preparação dos dados:

```
# Dados transformados para a ANOVA da RQ3
n = nrow(dados)

dados_ANOVA_RQ3 = data.frame(
  id = rep(dados$DM08_01, 2),
  tempo = rep(c("post-intervention", "after a delay"), each = n),
  score_teste_auto_controlado = c(dados$test2, dados$test3),
  precisao = c(dados$dist.post, dados$dist.delay),
  condicao = rep(dados$Condition, 2)
)

dados_ANOVA_RQ3$id = as.factor(dados_ANOVA_RQ3$id)
dados_ANOVA_RQ3$tempo = factor(dados_ANOVA_RQ3$tempo,
                               levels = c("post-intervention",
                                           "after a delay"))
dados_ANOVA_RQ3$condicao = as.factor(dados_ANOVA_RQ3$condicao)
```

Agora, partindo para a criação do modelo:

```
modelo_RQ3 = aov_ez(id = "id",
                    dv = "score_teste_auto_controlado",
                    between = "condicao",
                    within = "tempo",
                    data = dados_ANOVA_RQ3)
```

```
## Contrasts set to contr.sum for the following variables: condicao
```

```
# Output do modelo
modelo_RQ3
```

```
## Anova Table (Type 3 tests)
##
## Response: score_teste_auto_controlado
##      Effect    df  MSE    F ges p.value
## 1      condicao 1, 89 9.63 1.51 .011    .222
## 2        tempo 1, 89 5.27 2.35 .009    .129
## 3 condicao:tempo 1, 89 5.27 1.63 .006    .205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pelos *outputs* do modelo e considerando um valor para o **nível de significância** ($\alpha = 5\% = 0.05$), foram obtidos os seguintes resultados:

- A variável de **tempo** não exerceu um efeito significativo na performance de classificação dos testes, nos quais as ordens de estudo foram definidas pelos estudantes ($p = 0.129$):

- $F(1, 89) = 2.35$
- $p > \alpha \rightarrow$ aceitação da hipótese nula
- A variável de **condição** não exerceu um efeito significativo na performance de classificação dos testes, nos quais as ordens de estudo foram definidas pelos estudantes ($p = 0.222$):
 - $F(1, 89) = 1.51$
 - $p > \alpha \rightarrow$ aceitação da hipótese nula
- A interação das variáveis de **condição** e **tempo** (condição*tempo) não exerceu um efeito significativo na performance de classificação dos testes, nos quais as ordens de estudo foram definidas pelos estudantes ($p = 0.205$):
 - $F(1, 89) = 1.63$
 - $p > \alpha \rightarrow$ aceitação da hipótese nula

Figuras

RQ1. Como a crença dos participantes nas diferentes estratégias de estudo se comportou ao longo do tempo?

- Figura do autor

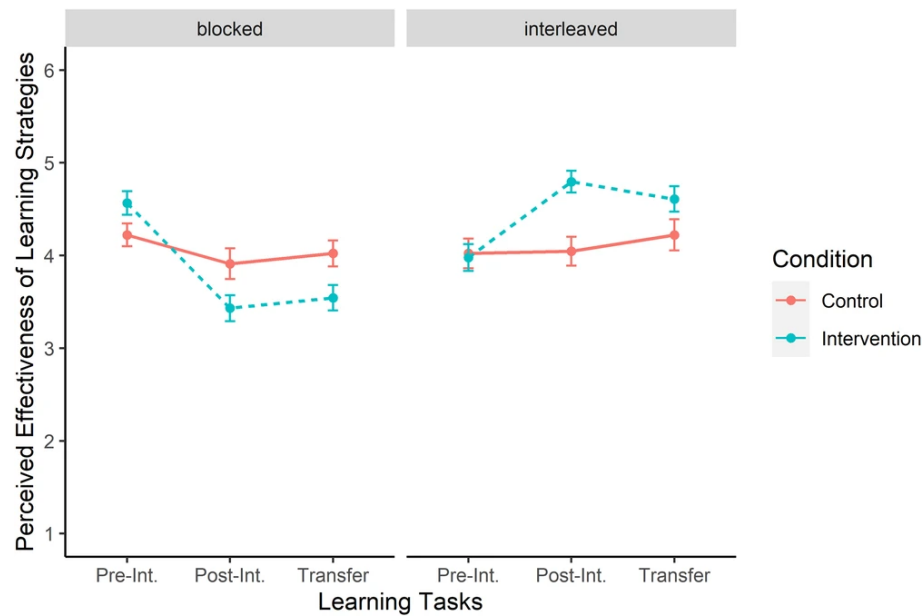
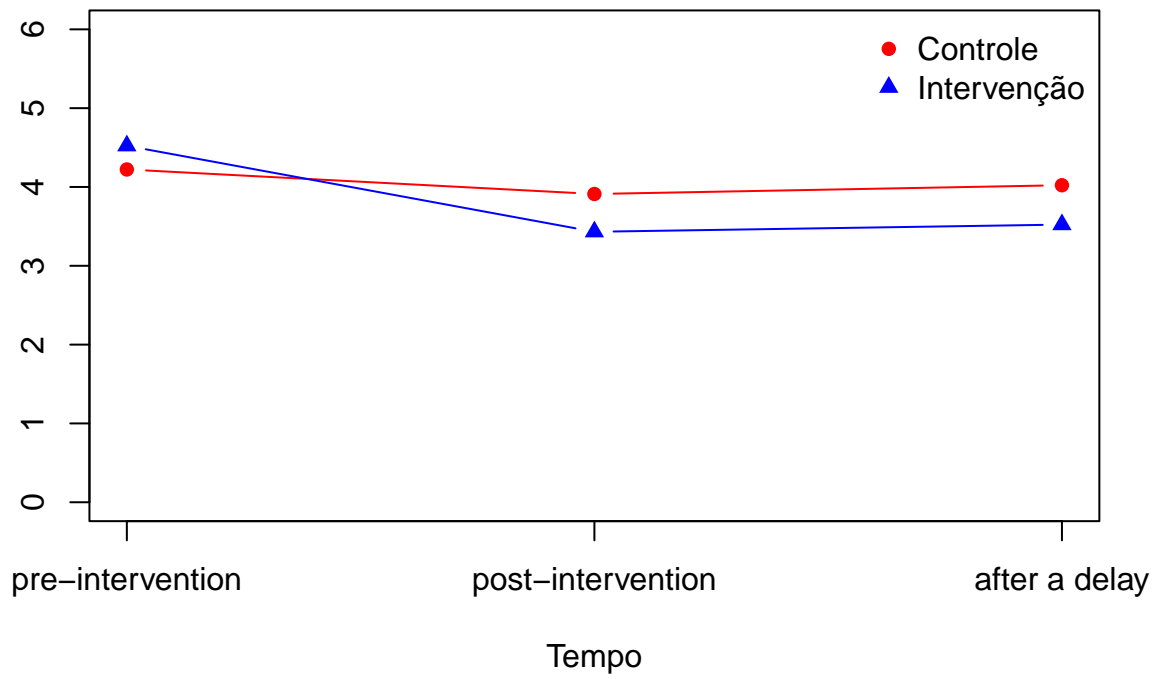


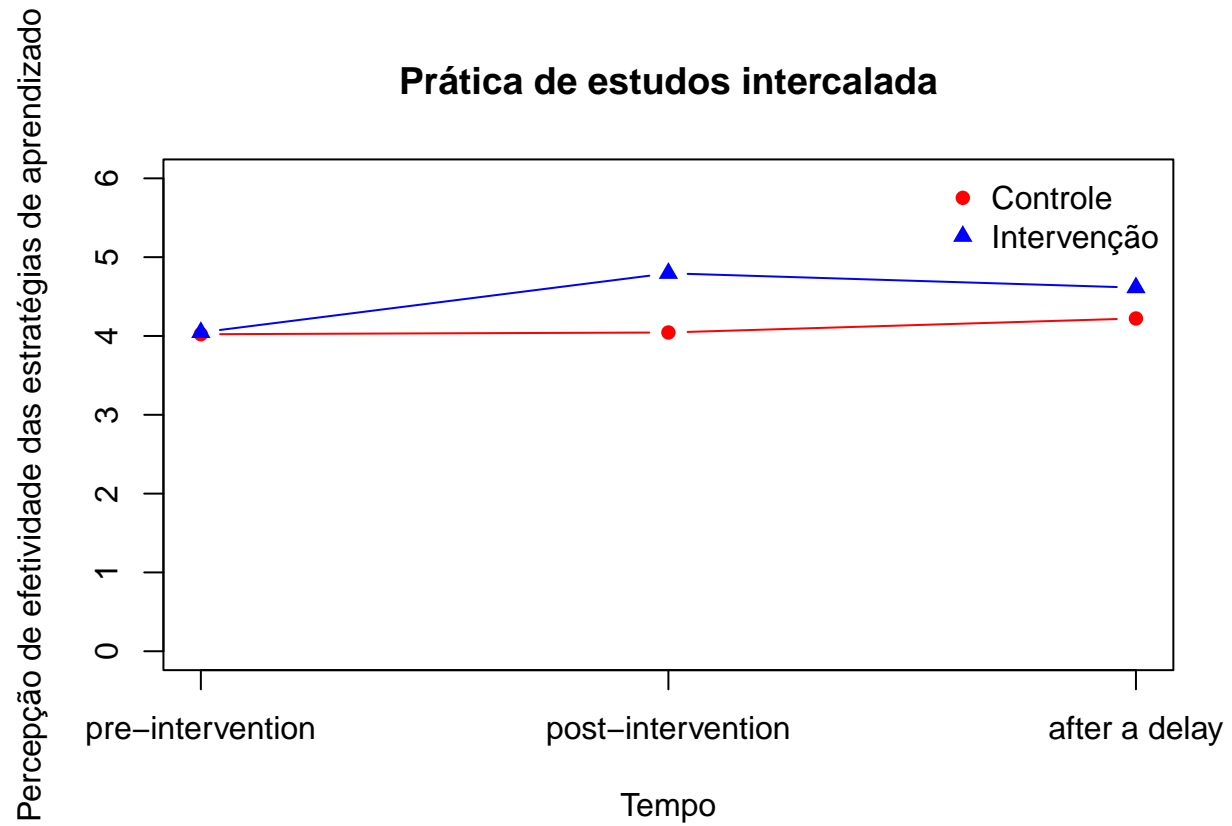
Figure 1: Perceived effectiveness of blocked and interleaved practice across learning tasks

- Figuras autorais

Percepção de efetividade das estratégias de aprendizado

Prática de estudos por blocos





RQ3. A prática de estudos intercalada possui algum efeito sobre a acurácia de classificação?

- Figura do autor

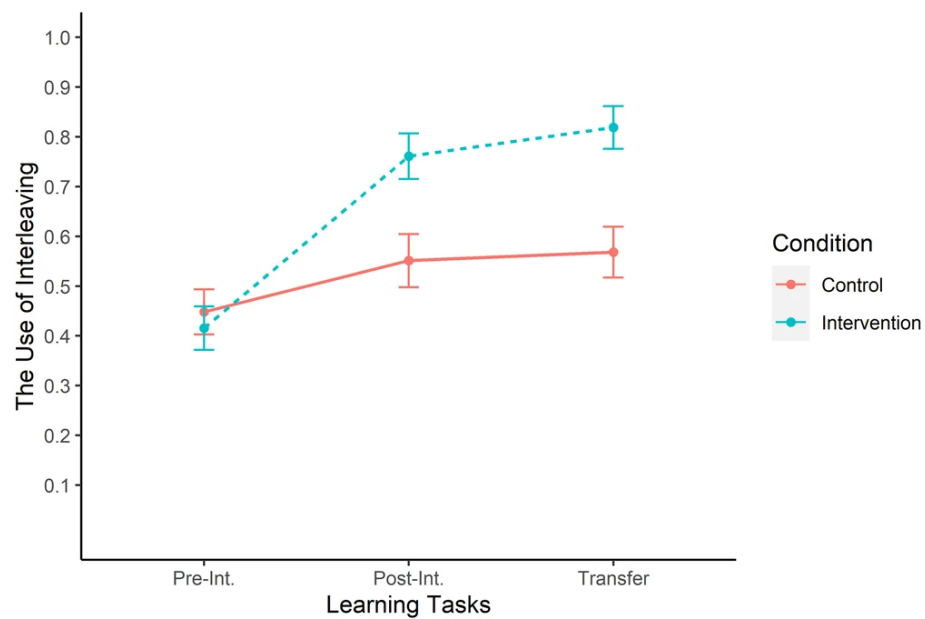
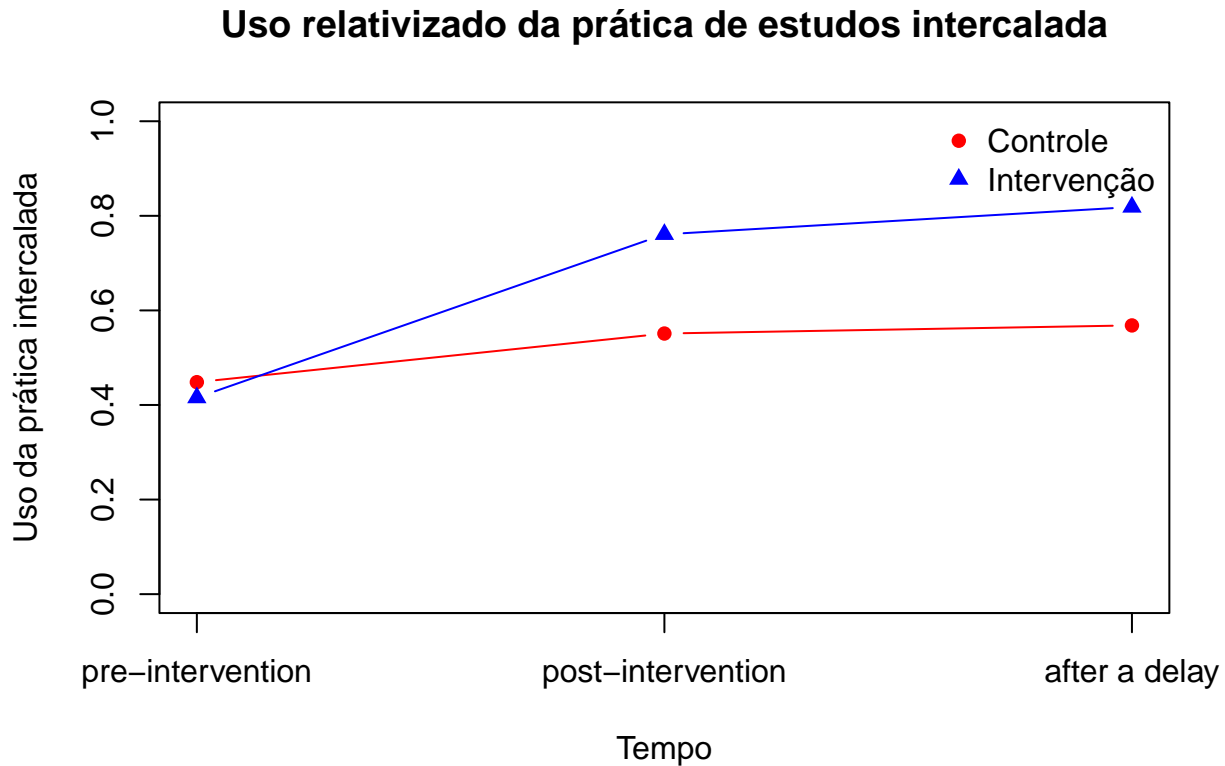


Figure 2: The use of interleaved practice (in proportions) across learning tasks

- Figura autoral



Conclusões

Através dessa pesquisa, é possível visualizar que existem diferentes formas de estudo, sendo que algumas se apresentam mais eficazes em detrimento de outras. Os autores apontam que a refutação foi uma estratégia conveniente do processo, por ter “forçado” os estudantes a questionarem-se e assim, entender de fato um determinado assunto.

Assim, torna-se viável que o método aplicado nessa pesquisa seja utilizado por universitários (ou estudantes, de um modo geral). É relevante mencionar que quando existe um interesse particular - ou o oposto - em uma categoria estudada, o comportamento dos estudantes frente aos estudos pode apresentar variações. Ademais, os autores sugerem que pesquisas futuras podem averiguar a eficácia da metodologia em situações cotidianas.

Contudo, vale destacar uma divergência metodológica relevante: enquanto os autores originais aplicaram testes t para comparar os grupos — testes que assumem a normalidade das distribuições —, verificamos que tal premissa não era atendida pelos dados. Em função disso, optamos por abordagens estatísticas não paramétricas, como os testes de permutação e o bootstrap, que são mais adequados em contextos onde não há garantia de normalidade. Apesar dessas diferenças na condução da análise estatística, as conclusões alcançadas permanecem consistentes.