**Summary Report**

# Definition

We measure the level of vocabulary learning by measuring the score achieved by individual participant.

Responsive Variable – score

Explanatory Variable – text coverage, genre, exam, occurrence

# Summary Statistics

> summary(df$text\_coverage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.000 3.000 4.000 3.968 5.000 6.000

> summary(df$genre)

Length Class Mode

1272 character character

> summary(df$exam)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.0 1.0 1.5 1.5 2.0 2.0

> summary(df$occurrence)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.0 1.0 1.5 1.5 2.0 2.0

> summary(df$score)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 1.000 2.000 2.245 3.000 8.000

# Plot

p1.pdf

p2.pdf

p3.pdf

# LM Model

Linear Regression Model:

Score ~ text coverage + genre + exam + occurrence

Fix effect: all variables

> summary(lm)

Call:

lm(formula = score ~ text\_coverage + genre + exam + occurrence,

data = df)

Residuals:

Min 1Q Median 3Q Max

-3.7965 -1.2959 -0.1017 1.1277 5.4771

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.40664 0.28120 22.783 < 2e-16 \*\*\*

text\_coverage -0.16685 0.03832 -4.354 1.44e-05 \*\*\*

genrenarrative 0.08464 0.10583 0.800 0.424

exam -0.68842 0.10583 -6.505 1.12e-10 \*\*\*

occurrence -1.67269 0.10583 -15.805 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.887 on 1267 degrees of freedom

Multiple R-squared: 0.1975, Adjusted R-squared: 0.195

F-statistic: 77.96 on 4 and 1267 DF, p-value: < 2.2e-16

Since genre’s p value > 0.05, we conclude that it is not statistically significant

> summary(lm)$coef

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.40664271 0.28120199 22.783063 1.411196e-96

text\_coverage -0.16684838 0.03831945 -4.354144 1.443544e-05

genrenarrative 0.08464332 0.10583326 0.799780 4.239882e-01

exam -0.68841691 0.10583326 -6.504731 1.117643e-10

occurrence -1.67269363 0.10583326 -15.804991 1.704997e-51

score = 6.40 – 0.16text\_coverage – 0.69exam – 1.67occurrence

# LMM Model

Linear Mixed Model:

Score ~ exam + occurrence + (1|text\_coverage)

Fix effect: exam, occurrence

Random intercept: text coverage

> # LMM model

> lmer1 <- lmer(score ~ exam + occurrence + (1|text\_coverage), data = df)

> summary(lmer1)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ exam + occurrence + (1 | text\_coverage)

Data: df

REML criterion at convergence: 5220.5

Scaled residuals:

Min 1Q Median 3Q Max

-1.94950 -0.64188 -0.04153 0.56380 2.80487

Random effects:

Groups Name Variance Std.Dev.

text\_coverage (Intercept) 0.1472 0.3836

Residual 3.5026 1.8715

Number of obs: 1272, groups: text\_coverage, 5

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.7684 0.2859 20.173

exam -0.6886 0.1050 -6.561

occurrence -1.6729 0.1050 -15.940

Correlation of Fixed Effects:

(Intr) exam

exam -0.551

occurrence -0.551 0.000

> ranef(lmer1)$text\_coverage

(Intercept)

2 0.21072928

3 0.15579188

4 0.03222413

5 0.24159668

6 -0.64034196

> coef(lmer1)$text\_coverage

(Intercept) exam occurrence

2 5.979157 -0.6885929 -1.67287

3 5.924220 -0.6885929 -1.67287

4 5.800652 -0.6885929 -1.67287

5 6.010025 -0.6885929 -1.67287

6 5.128086 -0.6885929 -1.67287

When text\_coverage = 2

score = 5.77 + 0.21text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 3

score = 5.77 + 0.15text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 4

score = 5.77 + 0.03text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 5

score = 5.77 + 0.24text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 6

score = 5.77 - 0.64text\_coverage - 0.68exam - 1.67occurrence

Conclusion:

* participant tend to learn fewer vocabulary when text\_coverage is higher
* the higher the text\_coverage, the lower the score achieved
* with the exception that when text\_coverage is 5, it has positive effect on score achieved
* participant tend the perform the worst when text\_coverage is 6

Score ~ text\_coverage + occurrence + (1|exam)

Fix effect: text\_coverage, occurrence

Random intercept: exam

> lmer2 <- lmer(score ~ text\_coverage + occurrence + (1|exam), data = df)

> summary(lmer2)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ text\_coverage + occurrence + (1 | exam)

Data: df

REML criterion at convergence: 5237.3

Scaled residuals:

Min 1Q Median 3Q Max

-1.98519 -0.66002 -0.03888 0.57949 2.87582

Random effects:

Groups Name Variance Std.Dev.

exam (Intercept) 0.2314 0.481

Residual 3.5608 1.887

Number of obs: 1272, groups: exam, 2

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.41642 0.40836 13.264

text\_coverage -0.16687 0.03831 -4.355

occurrence -1.67269 0.10582 -15.807

Correlation of Fixed Effects:

(Intr) txt\_cv

text\_coverg -0.372

occurrence -0.388 -0.001

> ranef(lmer2)$text\_coverage

NULL

> coef(lmer2)$text\_coverage

NULL

> lmer2 <- lmer(score ~ text\_coverage + occurrence + (1|exam), data = df)

> summary(lmer2)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ text\_coverage + occurrence + (1 | exam)

Data: df

REML criterion at convergence: 5237.3

Scaled residuals:

Min 1Q Median 3Q Max

-1.98519 -0.66002 -0.03888 0.57949 2.87582

Random effects:

Groups Name Variance Std.Dev.

exam (Intercept) 0.2314 0.481

Residual 3.5608 1.887

Number of obs: 1272, groups: exam, 2

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.41642 0.40836 13.264

text\_coverage -0.16687 0.03831 -4.355

occurrence -1.67269 0.10582 -15.807

Correlation of Fixed Effects:

(Intr) txt\_cv

text\_coverg -0.372

occurrence -0.388 -0.001

> ranef(lmer2)$exam

(Intercept)

1 0.3360757

2 -0.3360757

> coef(lmer2)$exam

(Intercept) text\_coverage occurrence

1 5.752497 -0.1668692 -1.672694

2 5.080346 -0.1668692 -1.672694

When exam = 1

score = 5.42 - 0.16text\_coverage + 0.33exam - 1.67occurrence

When exam = 2

score = 5.42 - 0.16text\_coverage - 0.33exam - 1.67occurrence

Conclusion:

* participant achieved higher score when taking the exam for the first time
* the second time the participant taking the exam, the lower score he/she achieved
* the higher exam time is, the lower the score the participant tend to achieve

Score ~ exam + text\_coverage + (1|occurrence)

Fix effect: exam, text\_coverage

Random intercept: occurrence

> lmer3 <- lmer(score ~ text\_coverage + exam + (1|occurrence), data = df)

> summary(lmer3)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ text\_coverage + exam + (1 | occurrence)

Data: df

REML criterion at convergence: 5239.1

Scaled residuals:

Min 1Q Median 3Q Max

-1.98772 -0.66255 -0.03871 0.58514 2.88190

Random effects:

Groups Name Variance Std.Dev.

occurrence (Intercept) 1.393 1.180

Residual 3.561 1.887

Number of obs: 1272, groups: occurrence, 2

Fixed effects:

Estimate Std. Error t value

(Intercept) 3.94000 0.86473 4.556

text\_coverage -0.16687 0.03831 -4.355

exam -0.68842 0.10582 -6.506

Correlation of Fixed Effects:

(Intr) txt\_cv

text\_coverg -0.176

exam -0.183 -0.001

> ranef(lmer3)$occurrence

(Intercept)

1 0.8329997

2 -0.8329997

> coef(lmer3)$occurrence

(Intercept) text\_coverage exam

1 4.772998 -0.1668672 -0.6884169

2 3.106999 -0.1668672 -0.6884169

When occurrence = 1

score = 3.94 - 0.16text\_coverage - 0.68exam + 0.83occurrence

When occurrence = 2

score = 3.94 - 0.16text\_coverage - 0.68exam – 0.83occurrence

Conclusion:

* participant achieved higher score when occurrence is 1
* the higher occurrence is, the lower the score the participant tend to achieve

Score ~ exam + occurrence + (1+ exam|text\_coverage)

Fix effect: exam, occurrence

Random intercept: text coverage

Random effect: exam

> summary(lmer4)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ exam + occurrence + (1 + exam | text\_coverage)

Data: df

REML criterion at convergence: 5220.4

Scaled residuals:

Min 1Q Median 3Q Max

-1.95108 -0.64137 -0.03685 0.56043 2.80390

Random effects:

Groups Name Variance Std.Dev. Corr

text\_coverage (Intercept) 0.1805332 0.42489

exam 0.0007528 0.02744 -1.00

Residual 3.5024956 1.87150

Number of obs: 1272, groups: text\_coverage, 5

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.7664 0.2974 19.390

exam -0.6872 0.1057 -6.504

occurrence -1.6729 0.1049 -15.940

Correlation of Fixed Effects:

(Intr) exam

exam -0.600

occurrence -0.529 0.000

optimizer (nloptwrap) convergence code: 0 (OK)

boundary (singular) fit: see help('isSingular')

> ranef(lmer4)$text\_coverage

(Intercept) exam

2 0.23278690 -0.015032505

3 0.17700376 -0.011430268

4 0.03843689 -0.002482127

5 0.26207534 -0.016923813

6 -0.71030289 0.045868713

> coef(lmer4)$text\_coverage

(Intercept) exam occurrence

2 5.999179 -0.7022699 -1.672874

3 5.943396 -0.6986677 -1.672874

4 5.804829 -0.6897196 -1.672874

5 6.028467 -0.7041612 -1.672874

6 5.056089 -0.6413687 -1.672874

When text\_coverage = 2

score = 5.77 + 0.23text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 3

score = 5.77 + 0.17text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 4

score = 5.77 + 0.04text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 5

score = 5.77 + 0.26text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 6

score = 5.77 - 0.71text\_coverage - 0.68exam - 1.67occurrence

Conclusion:

* similar conclusion as in 5.1.1
* when exam time is treated as random effect, it has small to no effects on level of vocabulary learning

Score ~ exam + occurrence + (1+ occurence|text\_coverage)

Fix effect: exam, occurrence

Random intercept: text coverage

Random effect: occurrence

> summary(lmer5)

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ exam + occurrence + (1 + occurrence | text\_coverage)

Data: df

REML criterion at convergence: 5216.8

Scaled residuals:

Min 1Q Median 3Q Max

-1.99145 -0.64337 -0.03106 0.59366 2.82303

Random effects:

Groups Name Variance Std.Dev. Corr

text\_coverage (Intercept) 0.55022 0.7418

occurrence 0.05659 0.2379 -1.00

Residual 3.49138 1.8685

Number of obs: 1272, groups: text\_coverage, 5

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.7509 0.4028 14.278

exam -0.6886 0.1048 -6.572

occurrence -1.6611 0.1493 -11.123

Correlation of Fixed Effects:

(Intr) exam

exam -0.390

occurrence -0.861 0.000

optimizer (nloptwrap) convergence code: 0 (OK)

boundary (singular) fit: see help('isSingular')

> ranef(lmer5)$text\_coverage

(Intercept) occurrence

2 0.42034799 -0.134810967

3 0.32477864 -0.104160655

4 0.01844935 -0.005916941

5 0.47095624 -0.151041679

6 -1.23453221 0.395930242

> coef(lmer5)

$text\_coverage

(Intercept) exam occurrence

2 6.171230 -0.6885772 -1.795960

3 6.075661 -0.6885772 -1.765310

4 5.769332 -0.6885772 -1.667066

5 6.221838 -0.6885772 -1.812191

6 4.516350 -0.6885772 -1.265219

attr(,"class")

[1] "coef.mer"

When text\_coverage = 2

score = 5.75 + 0.42text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 3

score = 5.75 + 0.32text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 4

score = 5.75 + 0.01text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 5

score = 5.75 + 0.47text\_coverage - 0.68exam - 1.67occurrence

When text\_coverage = 6

score = 5.75 - 1.23text\_coverage - 0.68exam - 1.67occurrence

Conclusion:

* when occurrence is treated as random effect, it has medium effect on level of vocabulary learning
* when occurrence is treated as random effect, it results participant’s score to a more variable range

# Summary Conclusion

* Genre has no significant effect on the level of vocabulary learned by the participant
* The more text coverage it is, the less the participant is going to learn(except when text coverage =5 for this experiment)
* Participant tend to learn the worst when text coverage is equal to 6
* Participant tend to score higher when taking the exam for the first time
* Participant tend to score higher when occurrence is lower
* No significant change of the level of learning is found when exam is treated as random effect
* Medium change of the level of learning is found when occurrence is treated as random effect, it increase the range of test score achieved by the participants