

MLM Assignment1

Alexandra Dacko Gaja Nenadovic Kyuri Park 5439043

05 June, 2022

Data Description

In this assignment, we analyze the ‘*exam*’ data, which carries the information about the pupils in English Schools. The data specifics are as follows:

- School: school id
- Student: student id
- Examscore (DV): exam score of each pupil
- LRTscore (IV): reading ability of each pupil (LRT, London Reading Test)
- AvsLRT (IV): average reading ability of each school

##	School	Student	Examscore	LRTscore	AvsLRT
## 1	1	1	0.26132	0.61906	0.16617
## 2	1	2	0.13407	0.20580	0.16617
## 3	1	3	-1.72390	-1.36460	0.16617
## 4	1	4	0.96759	0.20580	0.16617
## 5	1	5	0.54434	0.37110	0.16617
## 6	1	6	1.73490	2.18940	0.16617

1. Why should, in theory, a multilevel analysis be performed on these data?

The data has a naturally occurring hierarchical structure - pupils are nested within schools. In the data with a such nested structure, individual observations tend to be dependent. That is, students from the same school are more likely to be alike to each other. Correspondingly, the average correlation between pupils from the same school tends to be higher than the average correlation between pupils from different schools. Yet, the standard statistical models (e.g., single-level regression) rely much on the assumption of independence in the observations. Violating the independence assumption leads to biased standard error (i.e., underestimated SD) and spurious significant results. Therefore, here we should perform multilevel analysis, which can properly account for such dependencies in the data.

2. Provide descriptive statistics and check for outliers.

Table 1 shows the descriptive statistics for the sample of 4059 pupils in English Schools. We see the values of mean, median, minimum (min), maximum (max), standard deviation (sd), standard error (se), skewness, and kurtosis.

There are total 65 schools and the maximum number of students in a school is 198 while the minimum number of students in a school is 1. *Figure 1* shows the distribution of the number of students per school and we see that it varies a bit across the school (e.g., some schools have higher/fewer pupils), which is good to keep in mind.

Another thing to point out is that the mean/median value of **Examscore**, **LRTscore** is nearly 0 and the sd is 1. **AvsLRT** is distributed with mean around 0 as well, but has a smaller sd, which is equal to 0.3. As shown in *Figure 2*, **Examscore** and **LRTscore** seem to be evenly distributed following a symmetric distribution. This also aligns with the small value of skewness (i.e., $< |1|$) and kurtosis (i.e., < 1) for these variables. In addition, *Figure 2* shows that several univariate outliers exist in all three variables.

In *Figure 3*, we can check the bivariate outliers at each level (i.e., level 1 and level 2) as well as the linearity assumption. At level 1 (pupil), no influential outlier is observed and the linearity assumption seems to be satisfied. At level 2 (school), there seems to be a couple of data points that can be considered to be outliers (e.g, bottom right corner), and the linearity assumption appears to be met in level 2 as well.

Figure 4 shows the relationship between **Examscore** (on the y-axis) and **LRTscore**(on the x-axis) for the first and last 20 schools. We see that the intercepts and slopes for **LRTscore** vary considerably between the schools. Also, we see that students of different schools have different correlation coefficients. This could also serve as an evidence to account for the hierarchical structure of the data.

Note: the code for the figures can be found in the Appendix.

Table 1

Descriptive statistics (continued below)

	n	mean	sd	median	min	max
School	4059	31.01	18.94	29	1	65
Student	4059	38.7	30.26	33	1	198
Examscore	4059	-0.0001179	0.9989	0.004322	-3.666	3.666
LRTscore	4059	0.00181	0.9932	0.0405	-2.935	3.016
AvsLRT	4059	0.001811	0.3149	-0.0202	-0.756	0.6377

	skew	kurtosis	se
School	0.1979	-1.229	0.2973
Student	1.635	4.318	0.475
Examscore	0.003734	-0.02134	0.01568
LRTscore	-0.1276	0.06929	0.01559
AvsLRT	-0.203	-0.3467	0.004942

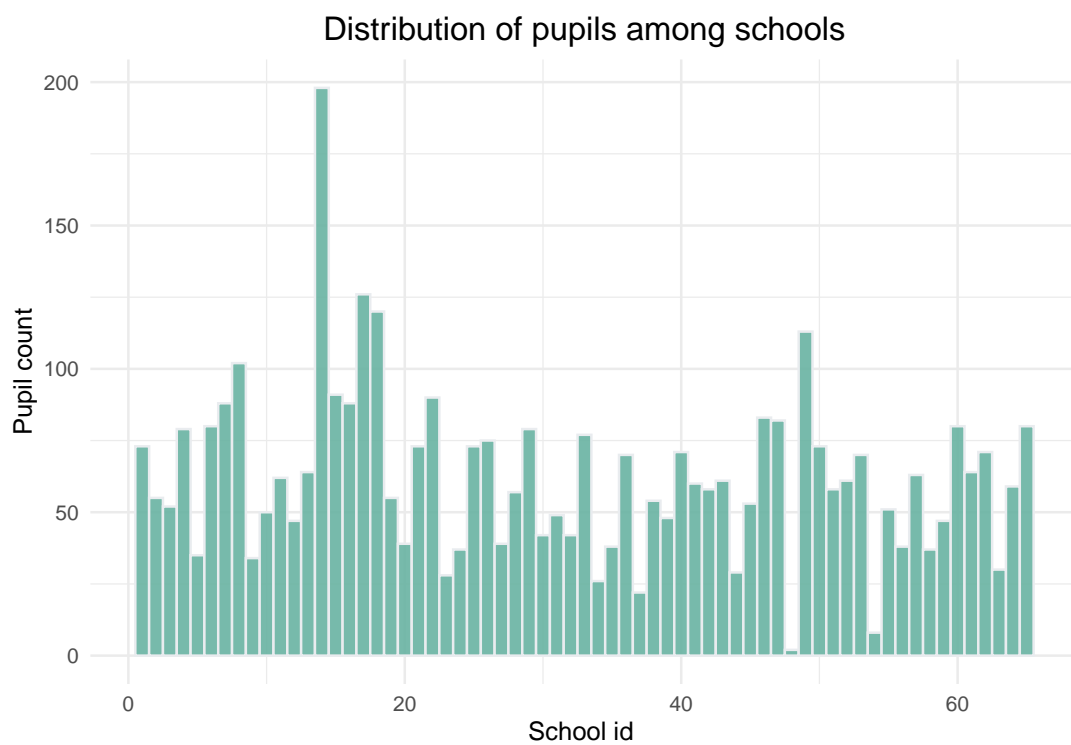


Figure 1: observation distribution among schools

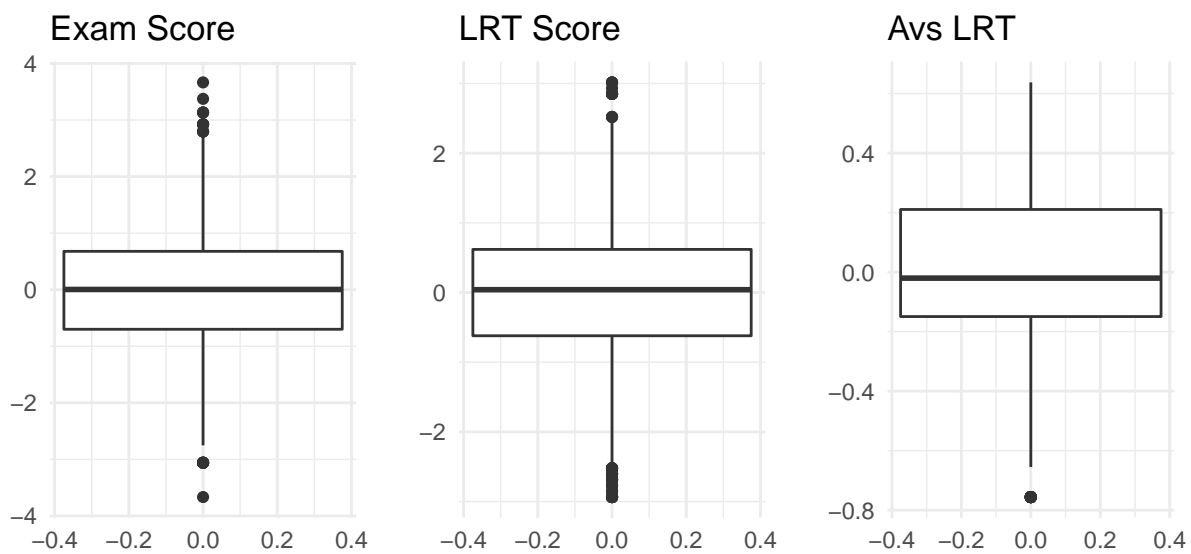


Figure 2: Distribution of Score variables

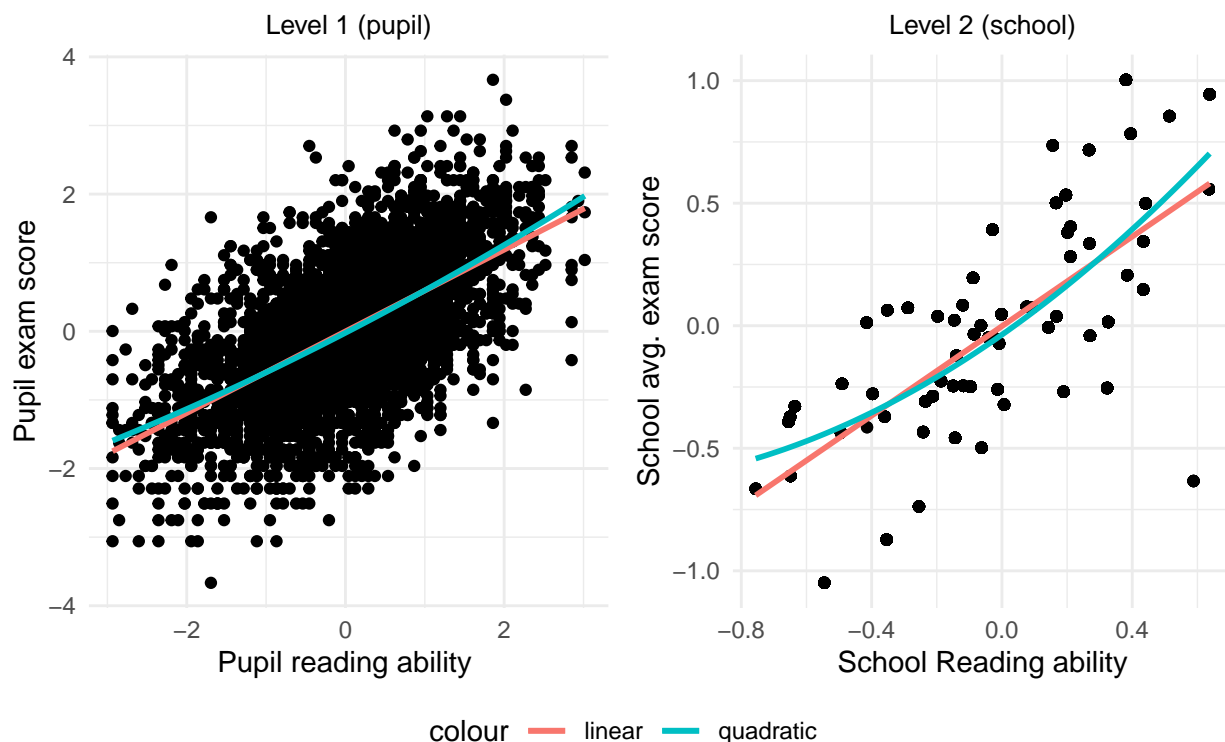


Figure 3: Scatterplots to inspect linearity and outliers

3. Answer the following questions

a. Should you perform a multilevel analysis or not?

Yes we should, because not only the data structure is clearly nested, but also the grouping structure (i.e., schools) accounts for about 17% of the total variance. In other words, the intraclass correlation – ICC: the proportion of the total variance explained by the grouping structure – is 0.166, which is deemed to be high enough that the multilevel analysis is warranted (the derivation of ICC can be found in 3e).

b. What are the null- and alternative hypotheses?

- $H_0: \sigma_{u0}^2 = 0$; Adding the random intercept term to the intercept-only model does not improve the fit of the model.
- $H_1: \sigma_{u0}^2 > 0$; Adding the random intercept term to the intercept-only model improves the fit of the model.

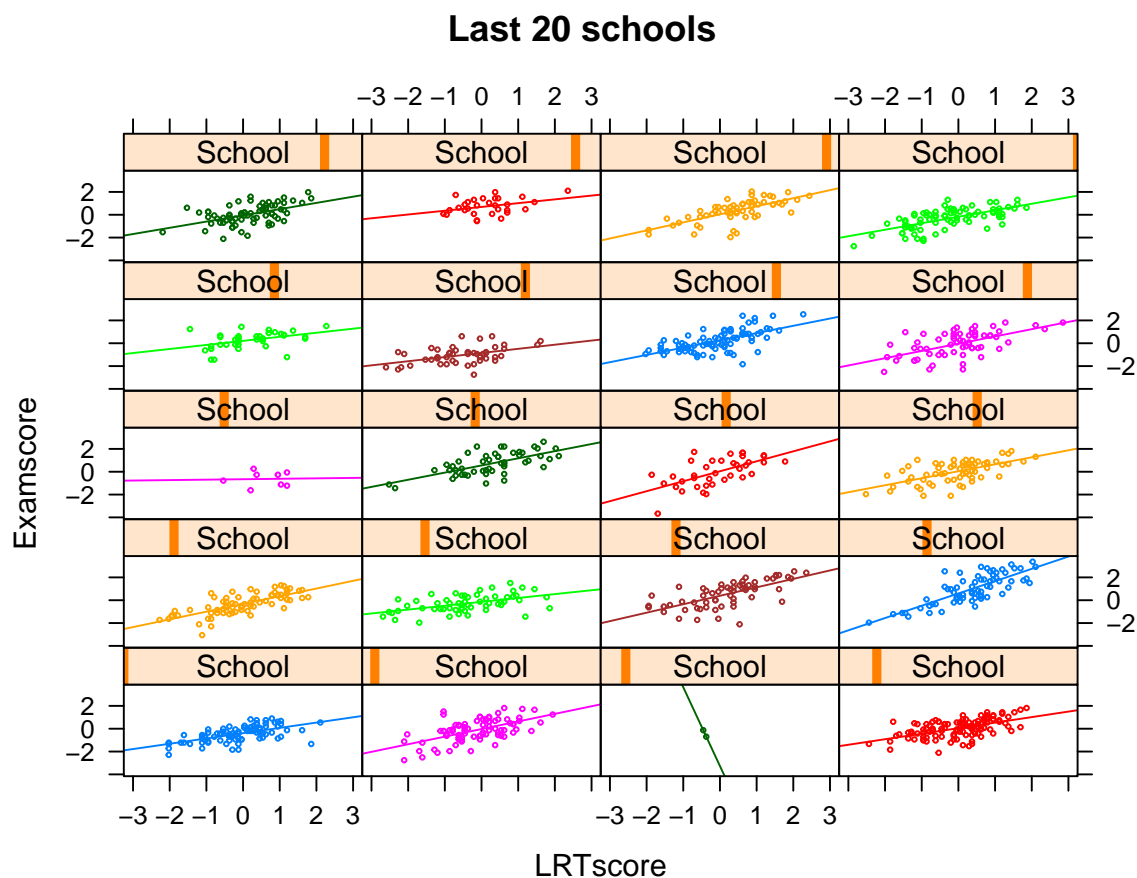
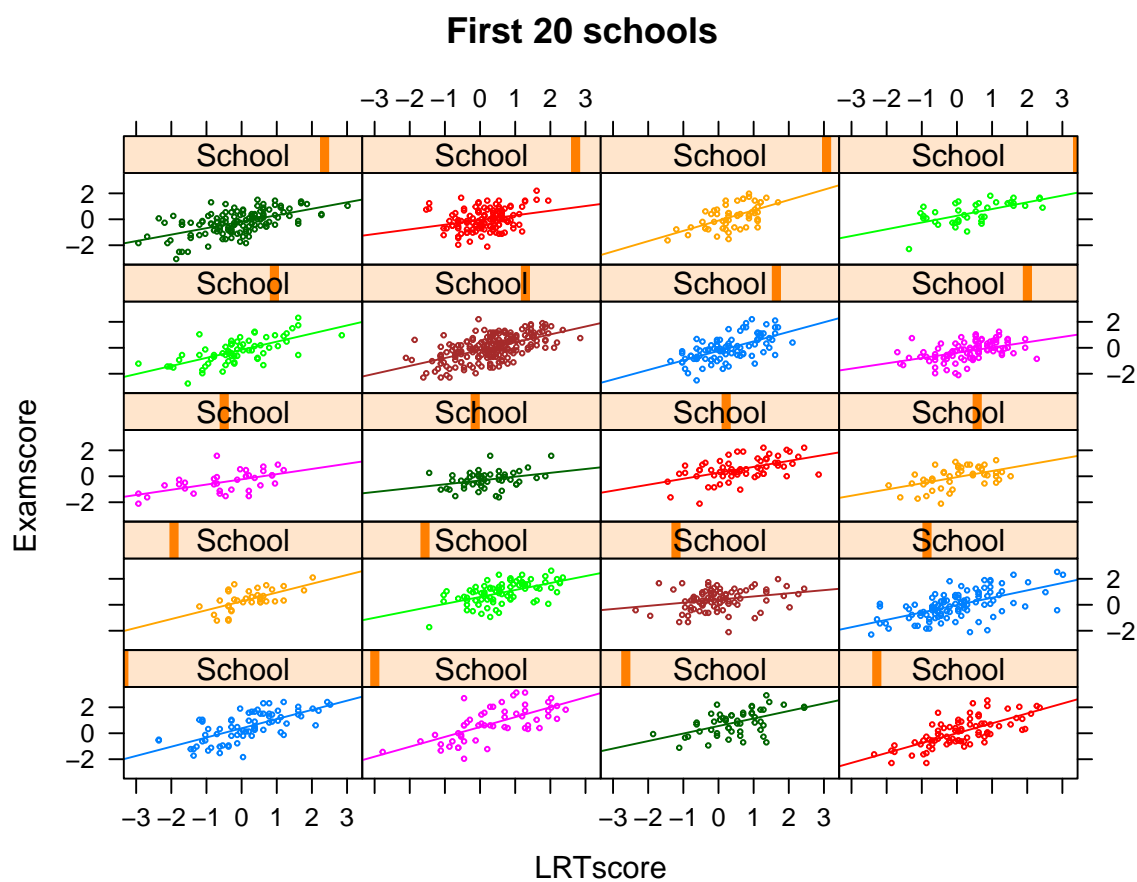


Figure 4: Scatterplots of each school to inspect the variabilities in intercept/slopes

c. What are the separate level 1 and 2 model equations, and the mixed model model equation?

- Level 1 Model Equation

$$y_{ij} = \beta_{0j} + e_{ij}$$

- y_{ij} refers to the exam scores (DV) of a pupil i in school j .
- β_{0j} refers to the intercept of the dependent variable in school j .
- e_{ij} refers to the residual error at an individual level (level 1).

- Level 2 Model Equation

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

- β_{0j} refers to the intercept of the dependent variable in school j .
- γ_{00} refers to the overall intercept, which is the grand mean of the exam scores (DV) across all the schools.
- e_{ij} refers to the residual error at an individual level (level 1).
- u_{0j} refers to the random residual error at the school level (level 2), which represents the deviation from the overall intercept (γ_{00}) of school j .

- Mixed Model Equation

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

d. Provide and interpret the relevant results.

As shown in *Table 3*, the intercept term (i.e., the average **Examscore** across all pupils) is 0.00 and -0.01 in the single-level model (M_0) and the intercept-only multilevel model (M_1) respectively, which is almost identical. But M_1 decomposes the variance term in a variance at level 1 (σ_e^2) and level 2 ($\sigma_{u_0}^2$).

The deviance for the intercept-only model turns out to be significantly smaller than that of the single-level model, $\chi^2(1) = 498.72$, $p < .001$, indicating that u_{0j} is significantly greater than 0. AIC value is also lower with the intercept-only model ($AIC_{M_1} = 11016.6$) compared to the single-level model ($AIC_{M_0} = 11513.0$), which aligns with the deviance difference test result.

```
# model 0 : single level model with an intercept
model0 <- lm(Examscore~1, data = dat)
summary(model0)

# model 1 : intercept only
model1 <- lmer(Examscore~1 + (1|School) , REML = FALSE, data = dat)
summary(model1)

# check the significance of random intercept
anova(model1, model0)
```

Table 3

Single-level model and Intercept-only model

Model	M_0 : single-level model	M_1 : intercept only
Fixed part	Coefficient(SE)	Coefficient(SE)
Intercept (γ_{00})	-0.00(.02)	-0.01(.05)
Random part		
σ_e^2	1.02	0.85
σ_{u0}^2		0.17
Deviance	11509.0	11010.6
AIC	11513.0	11016.
Deviance difference^{ab}		498.72***

Note

*** $p < .001$, ** $p < .01$ ^a p-value for χ^2 test is one-sided p-value.^b Here the deviance difference represents $dev_{M_3} - dev_{M_2}$.**e. What is the intraclass correlation?**

The intraclass correlation (ρ) is calculated as follows:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

As shown below, the intraclass correlation equals to 0.166 in this case, which is considered to be large enough to perform the multilevel analysis.

```
# compute ICC
ICC <- 0.1686/(0.1686 + 0.8478)
cat(paste0("ICC = ", round(ICC, 3)))
```

```
## ICC = 0.166
```

f. What is your conclusion considering the overall question regarding the necessity of performing a multilevel analysis?

We conclude that a multilevel analysis is necessary in this case for the following reasons:

- 1) As shown in *Figure 4*, there seems to be substantial variation different schools.
- 2) ICC (= 0.17) tells us that there is quite some proportion of variance explained by the grouping structure (schools).
- 3) The chi-square difference test comparing models M_0 and M_1 results in the significant outcome in favor of the intercept-only multilevel model (M_1), $\chi^2(1) = 498.72$, $p < .001$. This indicates that the M_1 fits the data significantly better than the single-level model (M_0).

4. Which variables are predictors of exam-score? Provide and interpret the relevant results and provide your overall conclusion.

In M_2 with the level 1 predictor (LRTscore), the reading ability is a statistically significant predictor of Examscore, $b_{LRTscore} = 0.56$, $t(4052) = 45.20$, $p < .01$. It means that with each point higher on the LRTscore,

the **Examscore** is expected to increase by 0.56. By adding this level 1 predictor, the first-level residual error variance (σ_e^2) goes down to 0.57 and the second-level variance (σ_{u0}^2) goes down to 0.09, as shown in *Table 4*. Based on this, we can compute the proportion of variance explained at the first and second level using the intercept-only model as our benchmark: $R_1^2 = .33$ and $R_2^2 = .45$ respectively.

In M_3 with the level 1 and level 2 predictors, both reading ability, $b_{LRTscore} = 0.6, t(3991) = 44.628, p < .01$, and school average reading ability, $b_{AvsLRT} = 0.36, t(65) = 3.25, p < .01$, are significant predictors of (average) exam score. It indicates that with each point higher on the **LRTscore**, the **Examscore** is expected to increase by 0.6, and for each point increase in the **AvsLRT**, the average **Examscore** for schools is expected to increase by 0.36. Adding the level 2 predictor to the model explains an additional part of the level 2 variance. As seen in *Table 4* σ_{u0}^2 goes down to 0.08 and this correspondingly produces $R_2^2 = .55$.

The significance of adding **LRTscore** to the intercept-only model (M_1):

The deviance of M_2 is significantly smaller than the deviance of M_1 , $\chi^2(1) = 1653.4, p < .001$ (see *Table 3* and *Table 4*). It indicates that M_2 fits significantly better than the intercept-only model M_1 . AIC also corresponds to this, as $AIC_{M_2} = 9365.2$ is lower than $AIC_{M_1} = 11016.6$.

The significance of adding **AvsLRT** to M_2 with the intercept and **LRTscore** predictor:

The deviance of M_3 is significantly smaller than the deviance of the M_2 , $\chi^2(1) = 9.6227, p < .001$ (see *Table 4*). It means that M_3 fits significantly better than M_2 . Again, AIC here is in accordance with the likelihood ratio test result, as $AIC_{M_3} = 9357.6$ is lower than $AIC_{M_2} = 9365.2$.

Therefore, it is concluded that M_3 with the level 1 and level 2 predictors is a better model and accordingly both **LRTscore** and **AvsLRT** are significant predictors of the **Examscore**.

```
# model2 : add level1 predictor
model2 <- lmer(Examscore ~ 1 + LRTscore + (1|School), REML=FALSE, data=dat)
summary(model2)

# model3 : add level2 predictor
model3 <- lmer(Examscore ~ 1 + LRTscore + AvsLRT + (1|School), REML=FALSE, data=dat)
summary(model3)

# compare the models: evaluate the significance of adding "LRTscore"
anova(model2, model1)

# compare the models: evaluate the significance of adding "AvsLRT"
anova(model3, model2)
```

5. Is the relation between reading-score and exam-score the same in all schools?

a. What are the null- and alternative hypotheses?

- $H_0: \sigma_{u1}^2 = 0$. The slope is equal across the schools.
- $H_1: \sigma_{u1}^2 > 0$. The slope varies across the schools.

Table 4

Model with level 1 predictor and level 1 and 2 predictors

Model	M_2 : level 1 predictor	M_3 : level 1 & 2
Fixed part	Coefficient(SE)	Coefficient(SE)
Intercept (γ_{00})	0.00 (.04)	0.01(.04)
LRTscore (γ_{10})	0.56(.01)***	0.56(.01)***
AvsLRT (γ_{01})		0.36(.11)**
Random part		
σ_e^2	0.57	0.57
σ_{u0}^2	0.09	0.08
Deviance	9357.2	9347.6
AIC	9365.2	9357.6
Deviance difference^{ab}		9.62***

Note

*** $p < .001$, ** $p < .01$ ^ap-value for χ^2 test is one-sided p-value.^b Here the deviance difference represents $dev_{M_3} - dev_{M_2}$

b. Provide and interpret the relevant results.

Again, both **LRTscore**, $b_{LRTscore} = 0.55, t(57) = 27.376, p < .01$, and **AvsLRT**, $b_{AvsLRT} = 0.29, t(66) = 2.792, p < .01$, are significant predictors of (average) **Examscore**. For each one point increase in **LRTscore**, the **Examscore** is expected to go up by 0.55. However, now that we have slopes varying across the schools, this coefficient (γ_{10}) refers to the expected value across all schools. And the interpretation of regression coefficient for **AvsLRT** stays the same: for each point increase in **AvsLRT**, the average **Examscore** for a school is expected to increase by 0.29.

Having introduced the random slopes, now we also have the covariance between the intercept and slope, σ_{u01} . As shown in *Table 5*, σ_{u01} is estimated to be 0.01, which can be interpreted as positive deviation to the overall intercept results in positive deviation to the overall slope in a magnitude of 0.01.

The variance component values (σ_e^2 and σ_{u0}^2) are quite similar between M_4 with the random slopes and M_3 without the random slopes (see *Table 4* and *Table 5*). The variance of the slope for **LRTscore** in M_4 seems quite small, $\sigma_{u1}^2 = .01$, but the difference in deviance between M_4 and M_3 turns out to be significant, $\chi^2(2) = 37.191, p < .001$. It indicates that M_4 with the random slope fits significantly better than M_3 . Additionally, $AIC_{M_4} = 9324.4$ is lower than $AIC_{M_3} = 9357.6$, which again suggests the better fit of M_4 .

One thing to note here is that we cannot directly assess the hypotheses with the test we use. Since there are two additional parameters (i.e., $\sigma_{u1}^2, \sigma_{u01}$) in the model with random slopes, significant χ^2 test does not necessarily reflect significance of slope variance alone. We could test on each parameter significance separately, but this somewhat complicates the interpretation. Given that variation in slopes is implied even in the case when only the covariance parameter is significant, we would proceed with our test inferring that the variation in slopes exist.

```
# model4 : add the random coefficient
```

```
model4 <- lmer(Examscore ~ 1 + LRTscore + AvsLRT + (LRTscore|School), REML=FALSE, data=dat)
summary(model4)
```

```
# compare the models: evaluate the significance of adding random slope
```

```
anova(model4, model3)
```

Table 5

Model with random slope and covariance of intercept and slope

Model	M_4 : random slope
<i>Fixed part</i>	
Intercept (γ_{00})	Coefficient(SE) 0.00 (.04)
LRTscore (γ_{10})	0.55(.02) ^{***}
AvsLRT (γ_{01})	0.29(.11) ^{**}
<i>Random part</i>	
σ_e^2	0.55
σ_{u0}^2	0.07
σ_{u1}^2	0.01
σ_{u01}	0.01
Deviance	9310.4
AIC	9324.4
Deviance difference^{ab}	37.19 ^{***}

Note

^{***} $p < .001$, ^{**} $p < .01$ ^a p-value for χ^2 test is one-sided p-value.^b Here the deviance difference represents $dev_{M_3} - dev_{M_4}$

c. Provide an overall conclusion.

In *Figure 4*, we observed quite some substantial variation in the slopes across the schools.

And in *5b* we actually showed that the slope variance for **LRTscore** (σ_{u1}^2) is significant given the likelihood ratio test (LRT) result, $\chi^2(2) = 37.191$, $p < .001$. In accordance with the LRT result, AIC of M_4 is also turned out to be lower than AIC of M_3 (see *Table 4* and *Table 5*).

Given these overall results, we decide that M_4 with the random slope is a better model than M_3 and correspondingly conclude that the relation between reading score (**LRTscore**) and **Examscore** differs across the schools.

6. Let's take a closer look at the regression coefficient of the reading-score obtained in the model under question 5.

a. Construct a 95% confidence interval for the regression coefficient of reading-score.

The 95% confidence interval for the regression coefficient of the reading score lies between 0.513 and 0.592 as shown below.

```
# regression coefficient of LRT
est.LRT <- 0.552391
# standard error of coefficient of LRT
se.LRT <- 0.020178
# lower & upper bound of 95% CI = est.LRT - Z*SE & est.LRT + Z*SE
lower.ci <- 0.552391 - 1.96*se.LRT
upper.ci <- 0.552391 + 1.96*se.LRT
# 95% confidence interval
ci95 <- c(round(lower.ci, 3), round(upper.ci, 3))
cat(paste("95% confidence interval =", "[", ci95[1], "-", ci95[2], "]""))
```

```
## 95% confidence interval = [ 0.513 - 0.592 ]
```

b. Construct the 95% predictive interval for the regression coefficients of reading-score.

The 95 % predictive interval for the reading-score coefficient lies between 0.313 and 0.792 as shown below.

```
# regression coefficient of LRT
est.LRT <- 0.552391
# standard deviation of coefficient of LRT
sd.LRT <- 0.1220
# lower & upper bound of 95% PI = est.LRT - Z*SD & est.LRT + Z*SD
lower.pi <- 0.552391 - 1.96*sd.LRT
upper.pi <- 0.552391 + 1.96*sd.LRT
# 95% predictive interval
pi95 <- c(round(lower.pi, 3), round(upper.pi, 3))
cat(paste("95% predictive interval =", "[", pi95[1], "-", pi95[2], "]"))
```

```
## 95% predictive interval = [ 0.313 - 0.792 ]
```

c. Explain the different meanings of 6a. and 6b.

- Confidence interval (CI) indicates the degree of uncertainty in the estimates of the unknown true population parameter. If we would repeat the sampling procedure infinite times and construct a CI each time, then a certain proportion of the confidence intervals (e.g., 95%) would contain the true population parameter value. Hence, the 95% CI from 6a can be seen as one of the numerous CIs assuming a large number of repeated samples from the population.
- As indicated in its name, predictive interval (PI) predicts a range of values that a future observation will fall within, given the current observations. More specifically, a PI is an interval associated with a to-be-observed value with a specified probability (e.g., 95%) of that observation lying within the interval. The 95% PI from 6b thus implies that the 95% of the regression coefficients of LRTscore are predicted to fall within the interval [0.313 - 0.792].
- One of the fundamental differences between confidence interval and prediction interval is what it concerns: confidence interval is concerned with the population parameter, while prediction interval is concerned with a future observation. As prediction interval needs to account for not only the uncertainty in estimation, but also the random variation of individual values, a prediction interval is always wider than a confidence interval.

7. Can average reading-score explain (a part of) the different relations between reading-score and exam-score in different schools? Provide and interpret the relevant results and provide your overall conclusion.

In M_5 , we added the cross-level interaction `LRTscore:AvsLRT` as a predictor. As shown in *Table 5* and *Table 6*, the estimates for the fixed coefficients are fairly similar in M_4 and M_5 except for the estimate regression coefficient for `AvsLRT`. And both `LRTscore` and `AvsLRT` remain as significant predictors of `Examscore` in M_5 ,

$b_{LRTscore} = 0.56, t(55.57) = 29.680, p < .001, b_{AvsLRT} = 0.37, t(65) = 3.421, p < .01$. The interpretation now is that with each point increase in LRTscore, Examscore is expected to go up by .56 when AvsLRT=0, and with each point increase in AvsLRT, the school average Examscore is expected to go up by .37 when LRTscore=0. The cross-level interaction term LRTscore:AvsLRT is also significant, $b_{LRTscore:AvsLRT} = 0.16, t(61) = 2.859, p < .01$. Which means that the effect of LRTscore on Examscore increases by .16 with every point increase in AvsLRT. In other words, the expected differences in the Examscore between the pupils with different reading ability are higher in the schools with higher AvsLRT score than in the schools with lower AvsLRT score (see Figure 5). Hence, school average reading ability (AvsLRT) moderates the relation between the individual reading ability and exam score.

Comparison of the two models (M_4 and M_5) reveals that the variance component for LRTscore (σ_{u1}^2) decreases a bit. It indicates that the average reading ability explains a part of the variation of the slopes for LRTscore: adding interaction term explains about 23% of the variation in the of the slopes for LRTscore, $R_{CL-interaction}^2 = .23$. As shown in Table 6, the deviance also goes down in M_5 , and the deviance difference between M_4 and M_5 turns out to be significant, $\chi^2(1) = 7.5172, p < .01$, meaning that M_5 fits significantly better than M_4 . Also, $AIC_{M_5} = 9318.9$ is lower than $AIC_{M_4} = 9324.4$, which again tells us that M_5 with the cross-level interaction term fits better.

Given these overall results, we conclude that the average reading score (AvsLRT) can explain a part of the different relations between reading score (LRTscore) and Examscore. And correspondingly, M_5 with the cross-level interaction term is the better model.

```
# model5 : add cross-level interaction
model5 <- lmer(Examscore ~ 1 + LRTscore + AvsLRT + LRTscore*AvsLRT+ (1 + LRTscore|School),
              REML=FALSE, data=dat)
summary(model5)
# compare the models: evaluate the significance of adding the cross-level interaction
anova(model5, model4)
```

Table 6

Model with random slope, covariance, and interaction

Model	M_5 : cross-level interaction
Fixed part	
Intercept (γ_{00})	-0.01 (.04)
LRTscore (γ_{10})	0.56(.02)***
AvsLRT (γ_{01})	0.37(.11)**
LRTscore:AvsLRT (γ_{11})	0.16(.06)**
Random part	
σ_e^2	0.55
σ_{u0}^2	0.07
σ_{u1}^2	0.01
σ_{u01}	0.01
Deviance	9302.9
AIC	9318.9
Deviance difference^{ab}	7.51**

Note

*** $p < .001$, ** $p < .01$

^a p-value for χ^2 test is one-sided p-value.

^b Here the deviance difference represents $dev_{M_4} - dev_{M_5}$

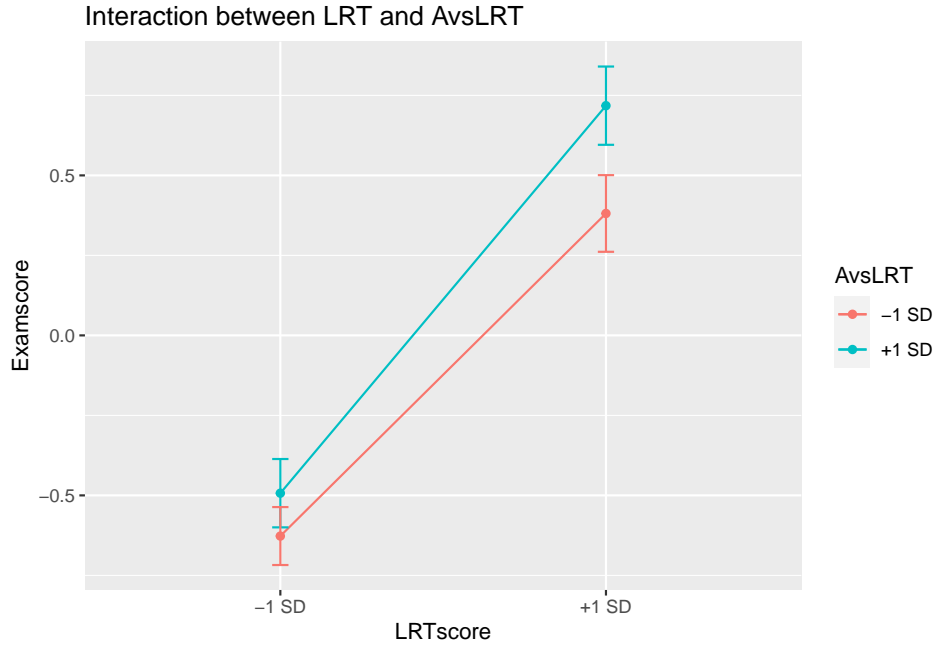


Figure 5: Interaction between LRT and AvsLRT

8. Choose a final model and provide the separate level 1 and 2 model equations, as well as the mixed model equation.

We chose **model 5** (full multilevel regression model) with the random intercepts, random slopes and cross-level interaction as our final model.

- Level 1 Model Equation

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$

- y_{ij} refers to the exam scores (DV) of a pupil i in school j .
- X_{ij} refers to the Level 1 predictor: reading ability of a pupil i in school j .
- β_{0j} refers to the intercept of the dependent variable in school j .
- β_{1j} refers to the slope coefficient for predictor X , which explains the relationship between X and DV.
- e_{ij} refers to the residual error at an individual level (level 1).

- Level 2 Model Equation

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

- Z_j refers to the Level 2 predictor: AvsLRT in school j .
- γ_{00} refers to the intercept to predict β_{0j} from Z_j .
- γ_{01} refers to the slope to predict β_{0j} from Z_j .
- γ_{10} refers to the intercept to predict β_{1j} from Z_j .
- γ_{11} refers to the slope to predict β_{1j} from Z_j .
- u_{0j} refers to the random residual error at the school level (level 2), which represents the deviation from the overall intercept (γ_{00}) of school j .

- u_{1j} refers to the random residual error at the school level (level 2), which represents the deviation from the overall slope (γ_{10}) of school j .

- Mixed Model Equation

$$y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

$$Examscore_{ij} = \gamma_{00} + \gamma_{10}LRT_{ij} + \gamma_{01}AvsLRT_j + \gamma_{11}AvsLRT_jLRT_{ij} + u_{0j} + u_{1j}LRT_{ij} + e_{ij}$$

9. Check the normality assumption for the level 1 residuals.

a. In R: The level 1 and 2 residuals can be accessed via the function `residuals(object, ...)`. Check the help file for more information and note the argument `level` and test what it does!

We have checked the help file and explored the `residuals()` function and it seems that the argument `level` does not actually give us access to the residuals at level 2. Therefore, we used the `ranef()` function to extract the group(school)-specific deviances to the intercept and slope.

```
# level 1 residuals
resid_lvl1<-residuals(model5)
# level 2 residuals
resid_lvl2<-ranef(model5)$School
```

b. Using the obtained residuals, (visually) inspect the normality assumption.

- Based on the *Figure 6(a)* where the residuals closely follow the straight diagonal line, it is concluded that the normality assumption is met in level 1 (e_{ij}).
- Given that we do not see any systematic pattern in *Figure 6 (b)*, it is concluded that the heteroscedasticity is met in level 1.
- *Figure 6(c)* shows that the the group-specific deviance to the intercept mostly follows a normal distribution. The lower tail slightly diverges, but it does not seem to be concerning. Hence, it is concluded that the normality assumption is met in level 2 intercept residuals.
- In *Figure 6(d)*, we see a single point deviating from the diagonal line (top right corner), while the rest seems to lie along the line. Despite the single deviating point, we conclude that normality assumption is not severely violated, given that the rest points follow the diagonal line fairly closely and overall they do not form a curve that deviates markedly from the line. In addition, just to note, according to Maas and Hox (2003), the non-normal residuals at the second level of the model have little or no effect on the parameter estimates. So a bit of deviation in our case should not be too concerning, when it comes to our analysis.

Contribution

- Aleksandra : coding and interpretation of the result.
 - Gaja : coding and interpretation of the result.
 - Kyuri: coding and interpretation of the result.
- Every group member equally contributes to the assignment. The teamwork was great.

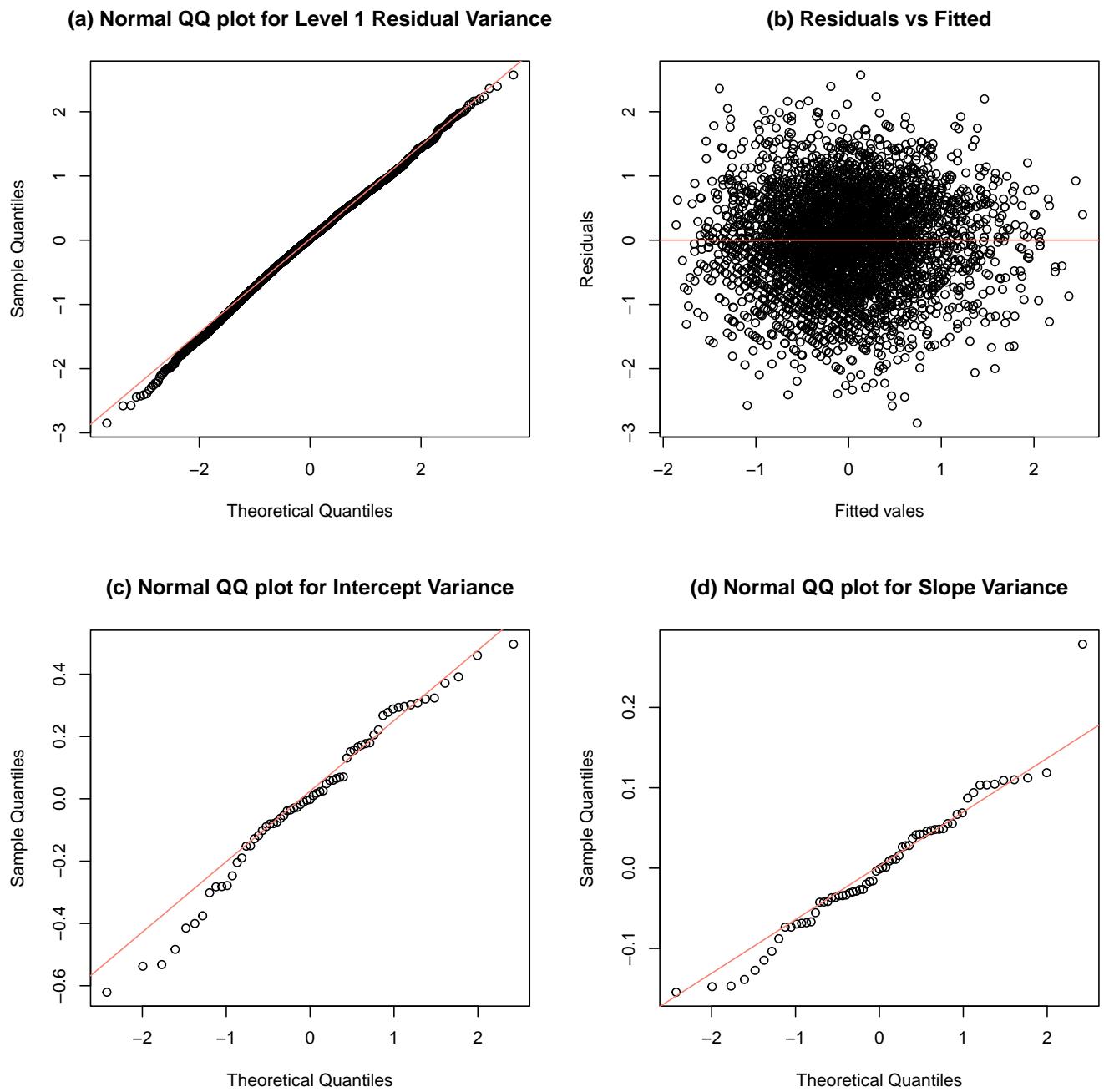


Figure 6: Q-Q plots for level 1 and level 2 residuals

Appendix

Code for Figure 1

```
dat %>%
  group_by(School) %>%
  ggplot(aes(x=School)) +
    geom_histogram( binwidth=1, fill="#69b3a2", color="#e9ecef", alpha=0.9) +
    theme_minimal() +
  xlab("School id ") +
  ylab("Pupil count")+
  ggtitle("Distribution of pupils among schools")+
  theme(plot.title = element_text(size = 14,hjust=0.5))
```

Code for Figure 2

```
box1 <- ggplot(dat, aes(y=Examscore)) + geom_boxplot() +
  theme_minimal() + labs(title = "Exam Score", y="")
box2 <- ggplot(dat, aes(y=LRTscore)) + geom_boxplot() +
  theme_minimal() + labs(title = "LRT Score", y = "")
box3 <- ggplot(dat, aes(y=AvsLRT)) + geom_boxplot() +
  theme_minimal() + labs(title = "Avs LRT", y = "")
ggarrange(box1, box2, box3, nrow=1)
```

Code for Figure 3

```
## check for linearity and outliers
# level1
p1 <- ggplot(dat,
  aes(x = LRTscore, y = Examscore)) +
  geom_point() +
  geom_smooth(method = "lm",
    aes(color = "linear"),
    se = FALSE) +
  geom_smooth(method = "lm",
    formula = y ~ x + I(x^2),
    aes(color = "quadratic"),
    se = FALSE) +
  theme_minimal() +
  xlab("Pupil reading ability ") +
  ylab("Pupil exam score")+
  ggtitle("Level 1 (pupil)")+
  theme(plot.title = element_text(size = 10,hjust=0.5))
```



```

# level2
p2 <- dat %>%
  group_by(School) %>%
  # aggregate the exam scores and store'em as : Examscore_aggr
  mutate(Examscore_aggr = mean(Examscore)) %>%
  ggplot(aes(x = AvsLRT, y = Examscore_aggr)) +
    geom_point() +
    geom_smooth(method = "lm",
               aes(color = "linear"),
               se = FALSE) +
    geom_smooth(method = "lm",
               formula = y ~ x + I(x^2),
               aes(color = "quadratic"),
               se = FALSE) +
    theme_minimal()+
    xlab("School Reading ability") +
    ylab("School avg. exam score")+
    ggtitle("Level 2 (school)")+
    theme(plot.title = element_text(size = 10, hjust=0.5))

ggarrange(p1, p2, nrow=1, common.legend = TRUE, legend="bottom")

```

Code for Figure 4

```

p3 <- xyplot(Examscore~LRTscore | School, groups= School, data=dat[dat$School %in% 1:20,],
             type=c('p','r'), auto.key=F, cex=0.3, main= "First 20 schools")

p4 <- xyplot(Examscore~LRTscore | School, groups= School, data=dat[dat$School %in% 46:65,],
             type=c('p','r'), auto.key=F, cex=0.3, main="Last 20 schools")

grid.arrange(p3, p4)

```

Code for Figure 5

```

# probing interaction
simple_slopes(model5)

graph_model(model5, y=Examscore, x=LRTscore, lines=AvsLRT,
             labels=list("title"="Interaction between LRT and AvsLRT"))

```

Code for Figure 6

```
# Set plot layout
layout(mat = matrix(c(1, 3, 2, 4), nrow = 2, ncol = 2), heights = c(2, 2), widths = c(2, 2))

# residual variance:  $e_{ij}$ 
qqnorm(resid_lvl1, main = "(a) Normal QQ plot for Level 1 Residual Variance")
qqline(resid_lvl1, col = "salmon")
# fitted vs resid( $e_{ij}$ ) plot: check heteroscedasticity
plot(fitted(model5), residuals(model5), main = "(b) Residuals vs Fitted",
     xlab="Fitted vales", ylab="Residuals")
abline(h=0, col="salmon")
# intercept variance:  $u_{0j}$ 
qqnorm(resid_lvl2[,1], main = "(c) Normal QQ plot for Intercept Variance")
qqline(resid_lvl2[,1], col = "salmon")
# slope variance:  $u_{1j}$ 
qqnorm(resid_lvl2[,2], main = "(d) Normal QQ plot for Slope Variance")
qqline(resid_lvl2[,2], col = "salmon")
```