EDUC 640

Introductory Statistics for Practitioners II

Week 9, Part 2

Interactions in Regression: Linear Mixed Models

Agenda

1. Example 1: Seeing the Mixed ANOVA from a prior lesson as a Mixed Linear Model
2. Example 2: Analyzing a comparative interrupted time series

Why do we say “Mixed”?

- “Mixed” in Mixed ANOVA refers to having both a **within**-person factor (e.g., a repeated measure) and a **between**-person factor (e.g., a group).

- A more accurate way to think about this is that mixed is about having **fixed effects** and **random effects**.

Fixed and Random Effects

Fixed Effects

- Our usual parameter estimates: **Intercepts** and **slopes**
- Represent the **average effect across all units** in the population
- These are what we interpret in our study: e.g., "**on average**, the intervention increases scores by 5 points."

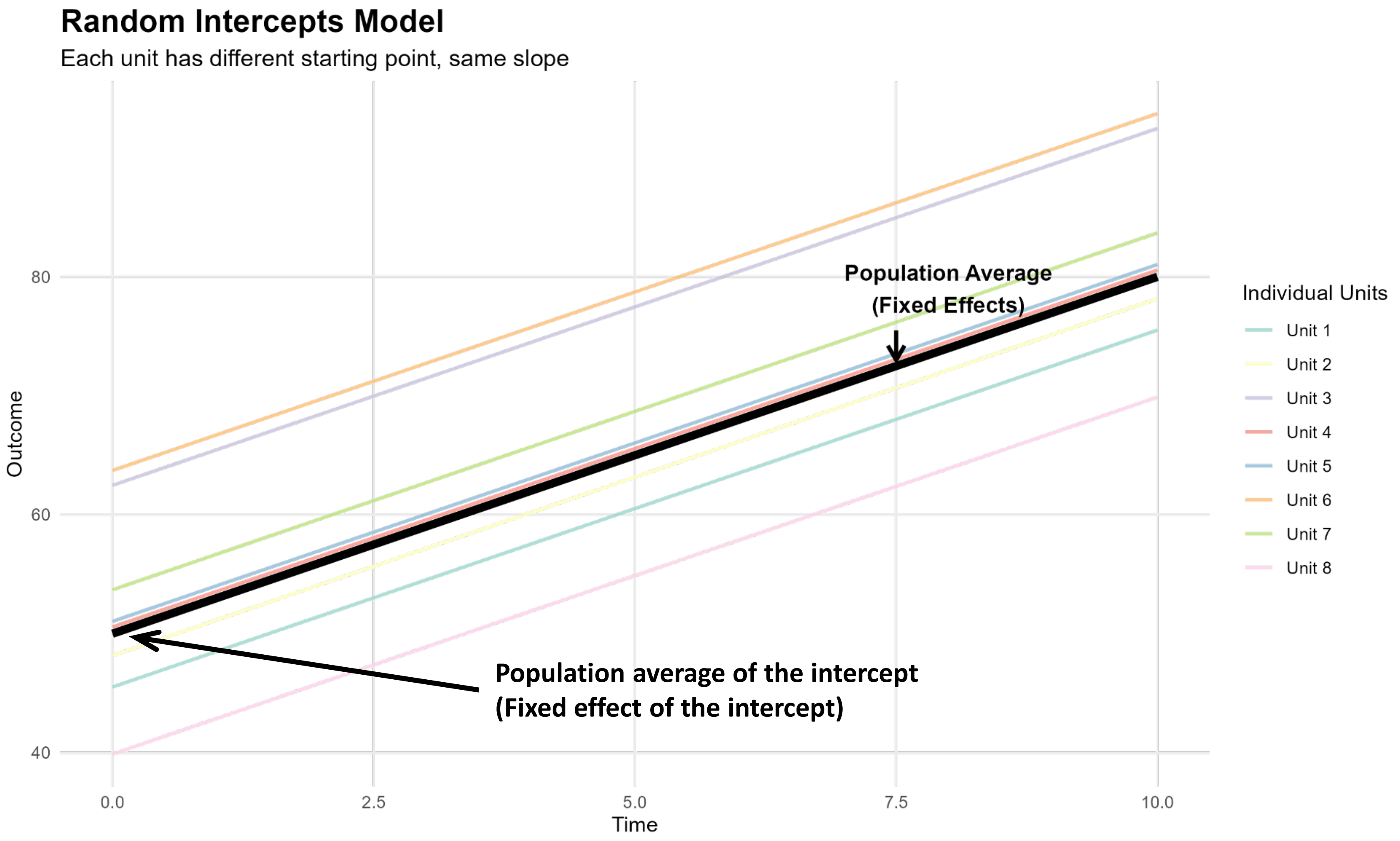
We interpret these.

Random Effects

- Allow individual **variation *among units*** (schools, students, etc.)
  - Recognizes that each unit *can have its own* intercept and/or slope
  - It's like adding "personalized adjustments" to the fixed effects.
- The model estimates the amount of variation. But, we do not interpret individual unit values.

We model these but do **not** interpret them.

Fixed and Random Effects



This plot is for demonstration. Plots like this can get busy. We usually only see the fixed effects line.

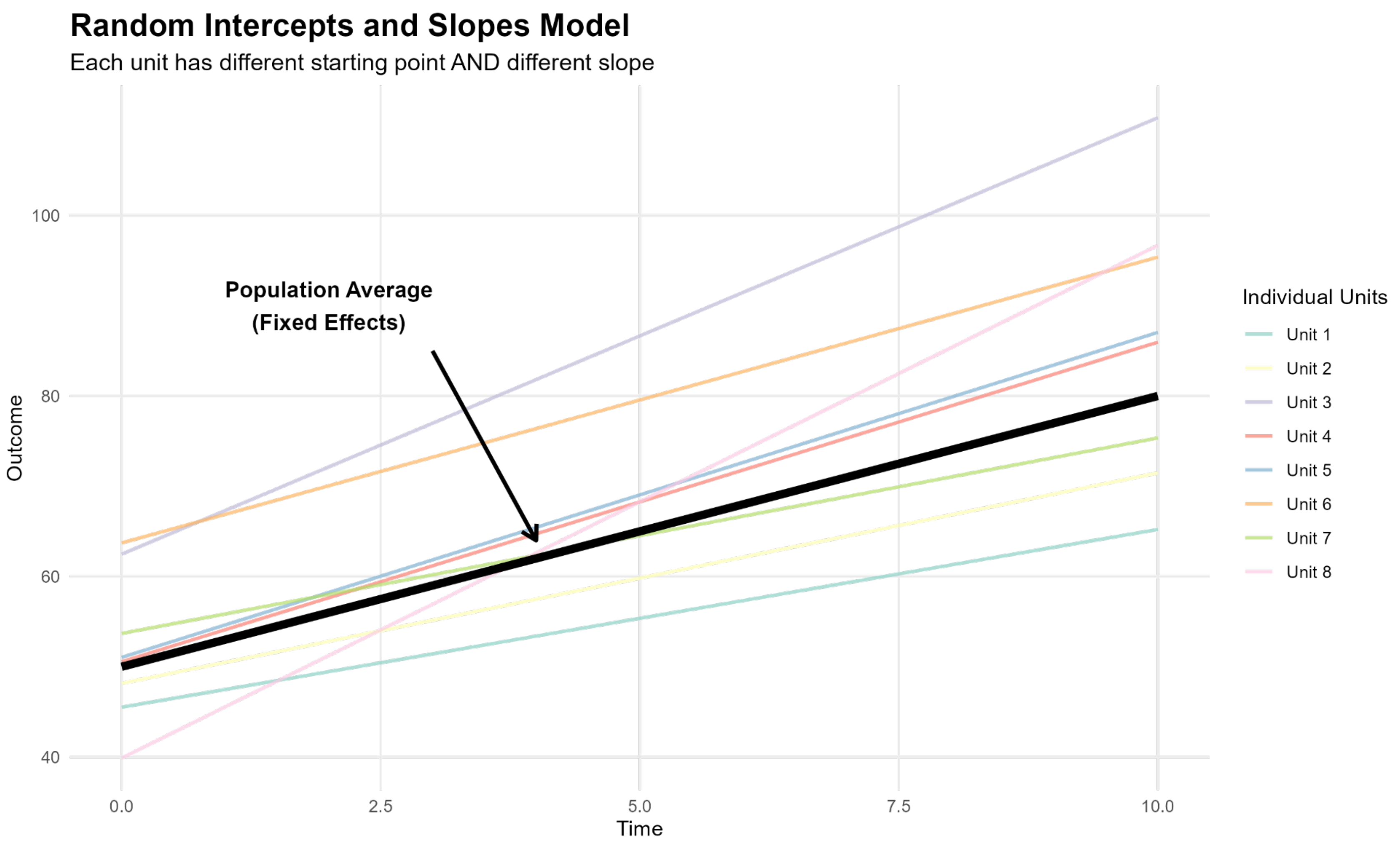
For example, maybe we're looking at students' growth over time on *Y*.

The growth rate, or slope, is estimated to be the same for all students. But where they start from can differ: Each student has their own intercept. This is the **random effect**.

The **fixed effect** of the intercept is the average score on *Y* when Time = 0.

The **fixed effect of the growth rate** is modeled but not the random variation among students in their growth rates.

Fixed and Random Effects



Students' starting levels on *Y* (the intercept) vary, and students' growth rates (the slope) vary. Those are **random effects** in the model.

The fixed effect of the intercept is the average starting point. **The fixed effect of the slope** is the average growth among students.

This plot is for demonstration. Plots like this can get busy. We usually only see the fixed effects line.

Terminology can vary across studies and disciplines

- Linear mixed models can go by other labels
  - Hierarchical linear models (AKA multilevel models)
  - Mixed effects models
  - Linear mixed-effects models
  - Random effects models

From Concept to Practice

Part 1: Familiar Territory

- Revisit our earlier mixed ANOVA example (the teacher bullying data set)
- Same data, same research question
- Show how GAMLj3 Linear Mixed Model gives identical results

☛ **Key insight:** Mixed ANOVA is actually a special case of Linear Mixed Models

Part 2: New Applications

- Consider a Comparative Interrupted Time Series (CITS) design
  - Reading intervention example with IEP vs. non-IEP students
- Use a Linear Mixed Model for the analysis

☛ **Key insight:** Some research questions require the flexibility of Linear Mixed Models

Part 1: In a prior lesson: We asked if there was a **difference between the two groups in their Time1-to-Time2 changes**.

H1: Teachers who participate in the workplace-social event series will become less intent to leave compared to their counterparts.

Data Set Format for Mixed ANOVA

Our data set included the ***repeated- measure levels as separate columns***. The dependent variable, which was intent-to-leave score, was recorded as the value within those columns.

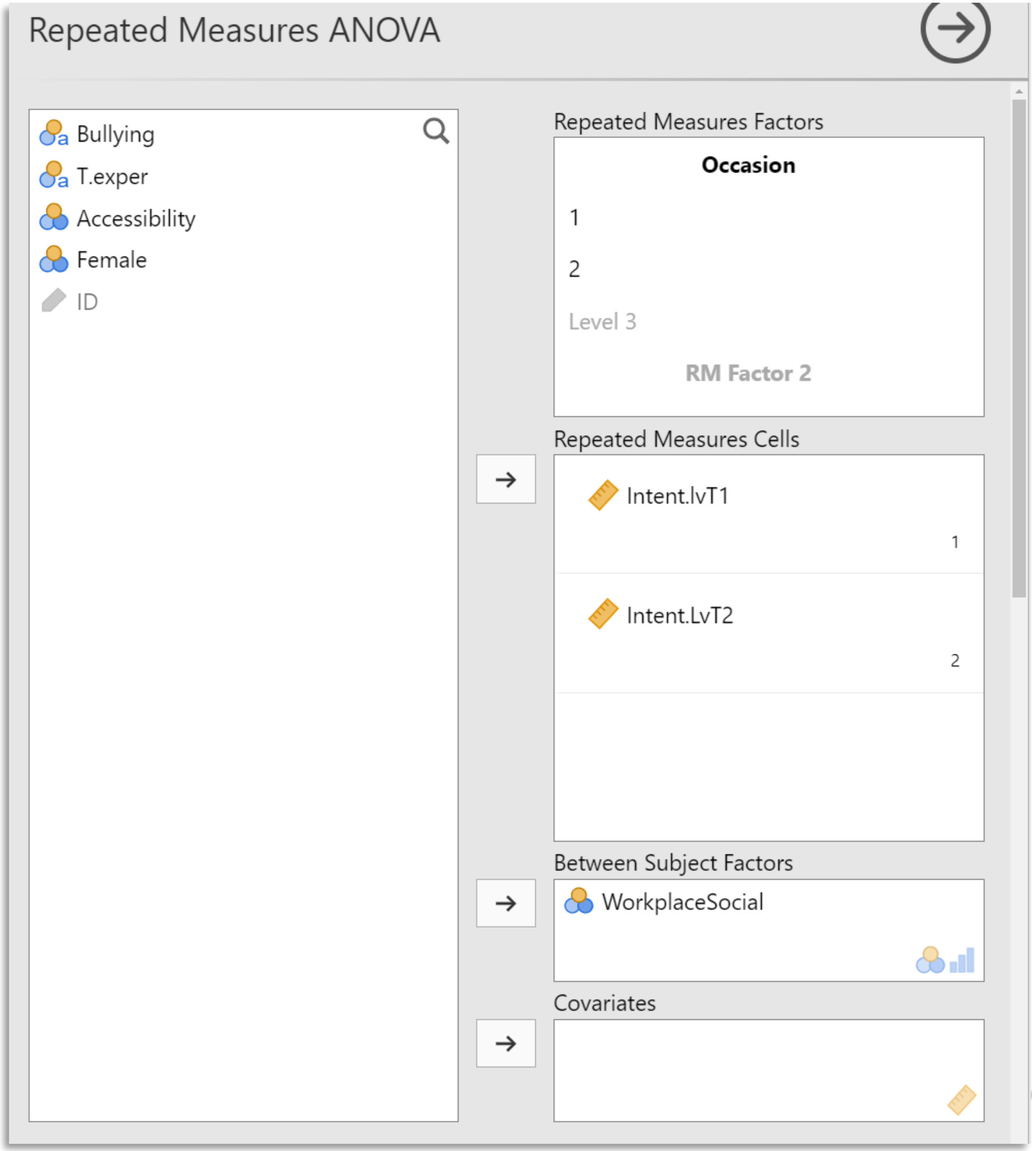
The **between-group variable**, Workplace-social event, was also a column. This indicated whether or not the person was offered this professional-development series.

For ANOVA, we use this **wide format** of the data.

	Repeated-measure Factor		Between-person Factor
ID	Intent.lvT1	Intent.LvT2	WorkplaceSocial
1001	6.9	16.6	No
1002	31.3	28.7	Yes
1003	18.4	18.2	No
1004	53.6	88.0	No
1005	30.4	26.8	Yes
1006	38.7	19.8	Yes
1007	53.6	90.1	No
1008	64.1	93.3	Yes
1009	0.9	8.3	No
1010	26.5	28.3	No
1011	28.6	27.7	No
1012	0.2	0.9	Yes

Mixed ANOVA Model Specification:

- Jamovi: ANOVA → Repeated Measures ANOVA
- Repeated factor specification
  - We give our repeated measure factor a name, here we typed “Occasion”.
  - Repeated measure labels are 1 and 2, for Time 1 and Time 2.
  - Drag the columns to their respective levels.
- Between-person factor
  - Add our group variable in the Between-Subject Factors field



First, we examined if the interaction effect was statistically significant.

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Occasion	5386.32	1	5386.32	24.87	< .001	0.14
Occasion ✻ WorkplaceSocial	5358.80	1	5358.80	24.75	< .001	0.14
Residual	34214.40	158	216.55			

Note. Type 3 Sums of Squares

[3]

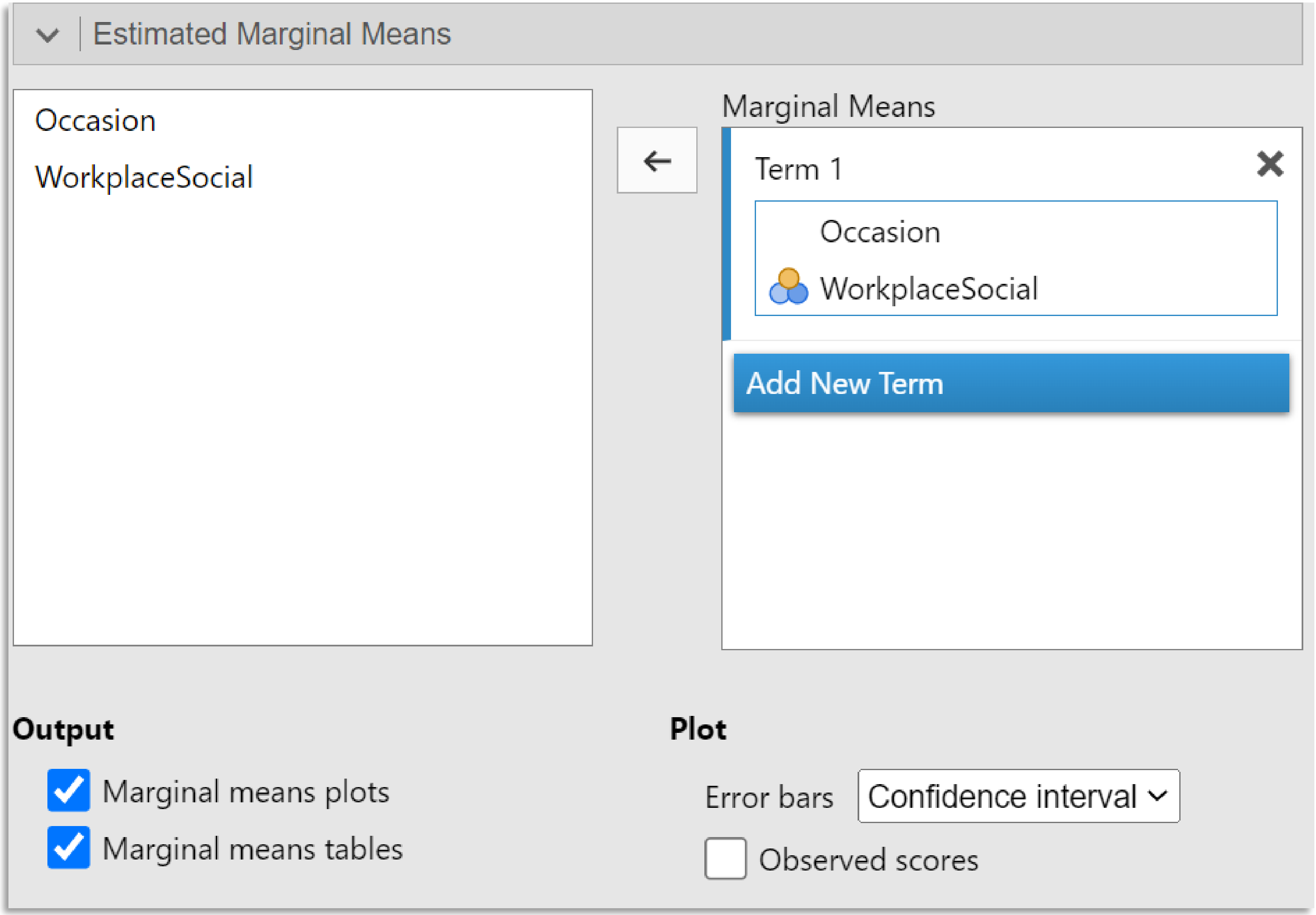
Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
WorkplaceSocial	3787.31	1	3787.31	2.71	0.102	0.02
Residual	221048.43	158	1399.04			

Note. Type 3 Sums of Squares

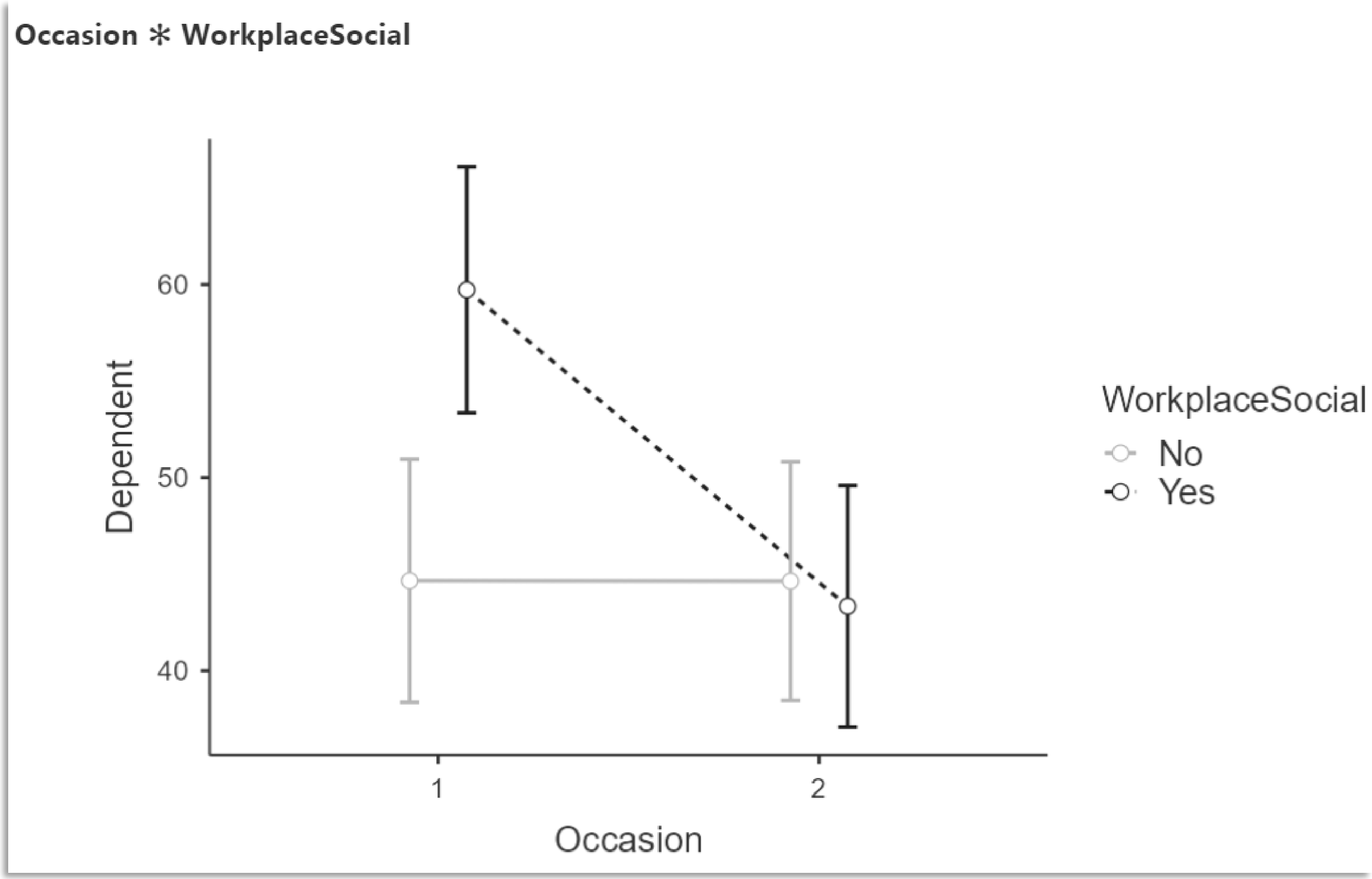
The interaction was significant, so we examined the pattern and tested the **simple effects**

- The estimated marginal means plot and/or table allows us to view the pattern.



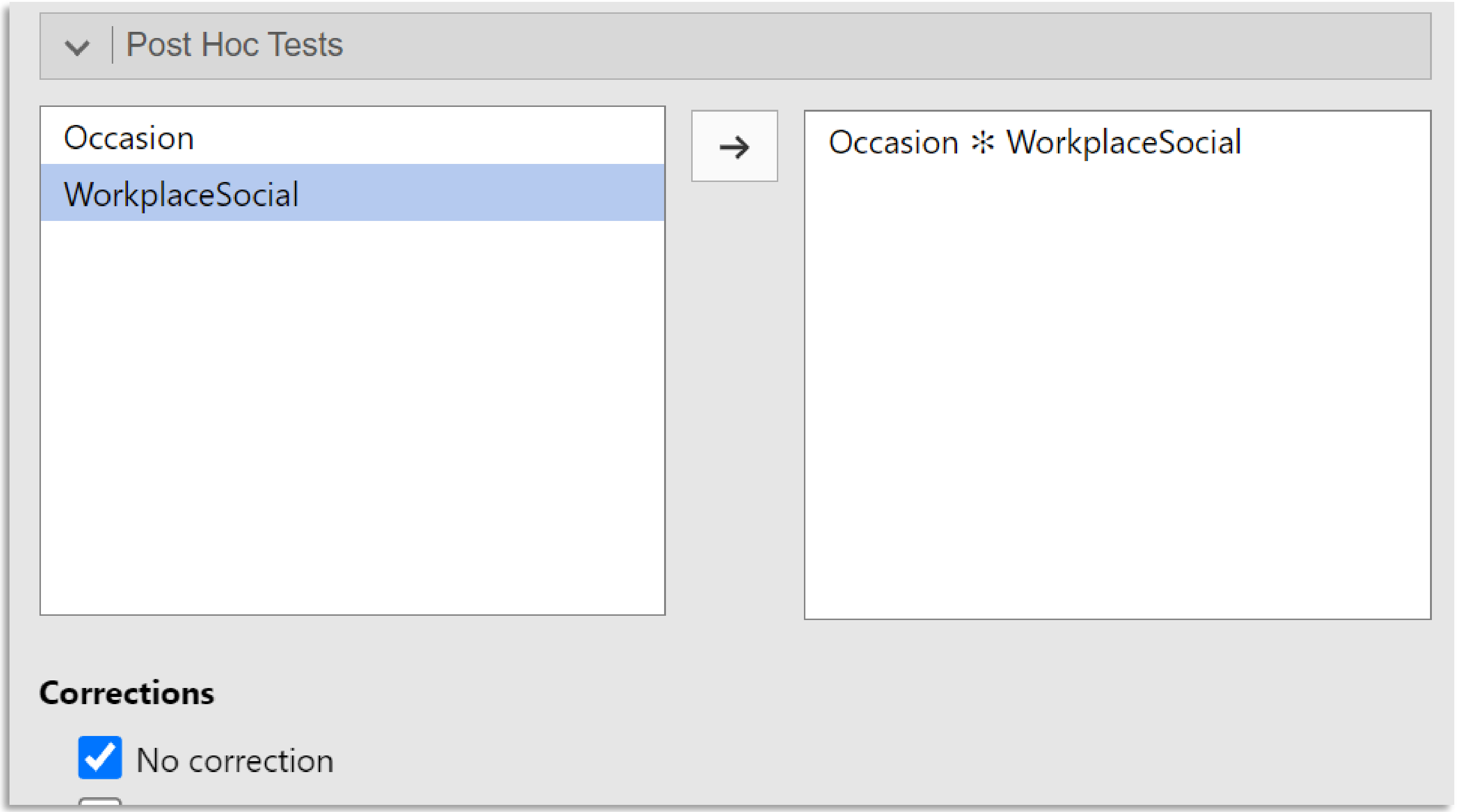
Interpret the Pattern and Test the Simple Effects

- We see a clear difference between the Yes and No groups in how their scores from Occasion 2 differed from Occasion 1.



Interpret the Pattern and **Test the Simple Effects**

- We can use the post-hoc tests menu. If we are examining all possible comparisons, we use one of the corrections. If not, we can focus on those comparisons of interest and use Bonferroni correct alpha or the adjusted Benjamini-Hochberg *p*-values.



Interpret the Pattern and **Test the Simple Effects**

Post Hoc Comparisons - Occasion ✻ WorkplaceSocial

Comparison									
Occasion	WorkplaceSocial		Occasion	WorkplaceSocial	Mean Difference	SE	df	t	p
1	No	-	1	Yes	−15.07	4.54	158.00	−3.32	0.001
		-	2	No	0.02	2.31	158.00	0.01	0.993
		-	2	Yes	1.33	4.49	158.00	0.29	0.768
	Yes	-	2	No	15.09	4.49	158.00	3.36	< .001
		-	2	Yes	16.39	2.34	158.00	7.00	< .001
2	No	-	2	Yes	1.30	4.45	158.00	0.29	0.770

- We are interested in two comparisons:
  1. Occasion 1 to 2 **within the Yes** group and
  2. Occasion 1 to 2 within the No group.

Interpret the Pattern and **Test the Simple Effects**

Post Hoc Comparisons - Occasion ✻ WorkplaceSocial

Comparison									
Occasion	WorkplaceSocial		Occasion	WorkplaceSocial	Mean Difference	SE	df	t	p
1	No	-	1	Yes	−15.07	4.54	158.00	−3.32	0.001
		-	2	No	0.02	2.31	158.00	0.01	0.993
		-	2	Yes	1.33	4.49	158.00	0.29	0.768
	Yes	-	2	No	15.09	4.49	158.00	3.36	< .001
		-	2	Yes	16.39	2.34	158.00	7.00	< .001
2	No	-	2	Yes	1.30	4.45	158.00	0.29	0.770

- We are interested in two comparisons:
  1. Occasion 1 to 2 within the Yes group and
  2. Occasion 1 to 2 **within the No** group.

Mixed ANOVA Results Writetup

To examine whether there was an association between the workplace-social intervention and a reduction in teachers' intent to leave, we conducted a mixed ANOVA examining the interaction between time (Time 1 and 2) and group (workplace-social vs. control). The interaction was statistically significant, $F(1, 158) = 24.75$, $p < .001$, with a large effect size (partial $\eta^2 = .135$). We set alpha at .025 to correct for multiple comparisons when examining the two simple effects of interest. For teachers in the workplace-social group, their 16.39-point decrease in intent-to-leave was statistically significant, $t(158) = 7.00$, $p < .001$. For teachers in the control group, the change in intent-to-leave was negligible and not statistically significant, $t(158) = 0.001$, $p = .993$.

Introducing a Linear Mixed Model: Data Format

Repeated Measures ANOVA: use **Wide Format**

ID	Intent.lvT1	Intent.LvT2	WorkplaceSocial
1001	6.9	16.6	No
1002	31.3	28.7	Yes
1003	18.4	18.2	No
1004	53.6	88.0	No
1005	30.4	26.8	Yes
1006	38.7	19.8	Yes
1007	53.6	90.1	No
1008	64.1	93.3	Yes
1009	0.9	8.3	No
1010	26.5	28.3	No
1011	28.6	27.7	No
1012	0.2	0.9	Yes

Linear Mixed Model: use **Long Format**

ID	Occasion	Intent.lv	WorkplaceSocial
1001	1	6.9	No
1001	2	16.6	No
1002	1	31.3	Yes
1002	2	28.7	Yes
1003	1	18.4	No
1003	2	18.2	No
1004	1	53.6	No
1004	2	88.0	No
1005	1	30.4	Yes
1005	2	26.8	Yes
1006	1	38.7	Yes

Data Layout in a Linear Mixed Model

Data Set Format for Linear Mixed Model

ID is a categorical variable (a factor). Each person is represented by multiple rows, one for each occasion.

Occasion is a factor variable, with 1 and 2 as its levels in this analysis.

The dependent variable, intent-to-leave score is a continuous variable.

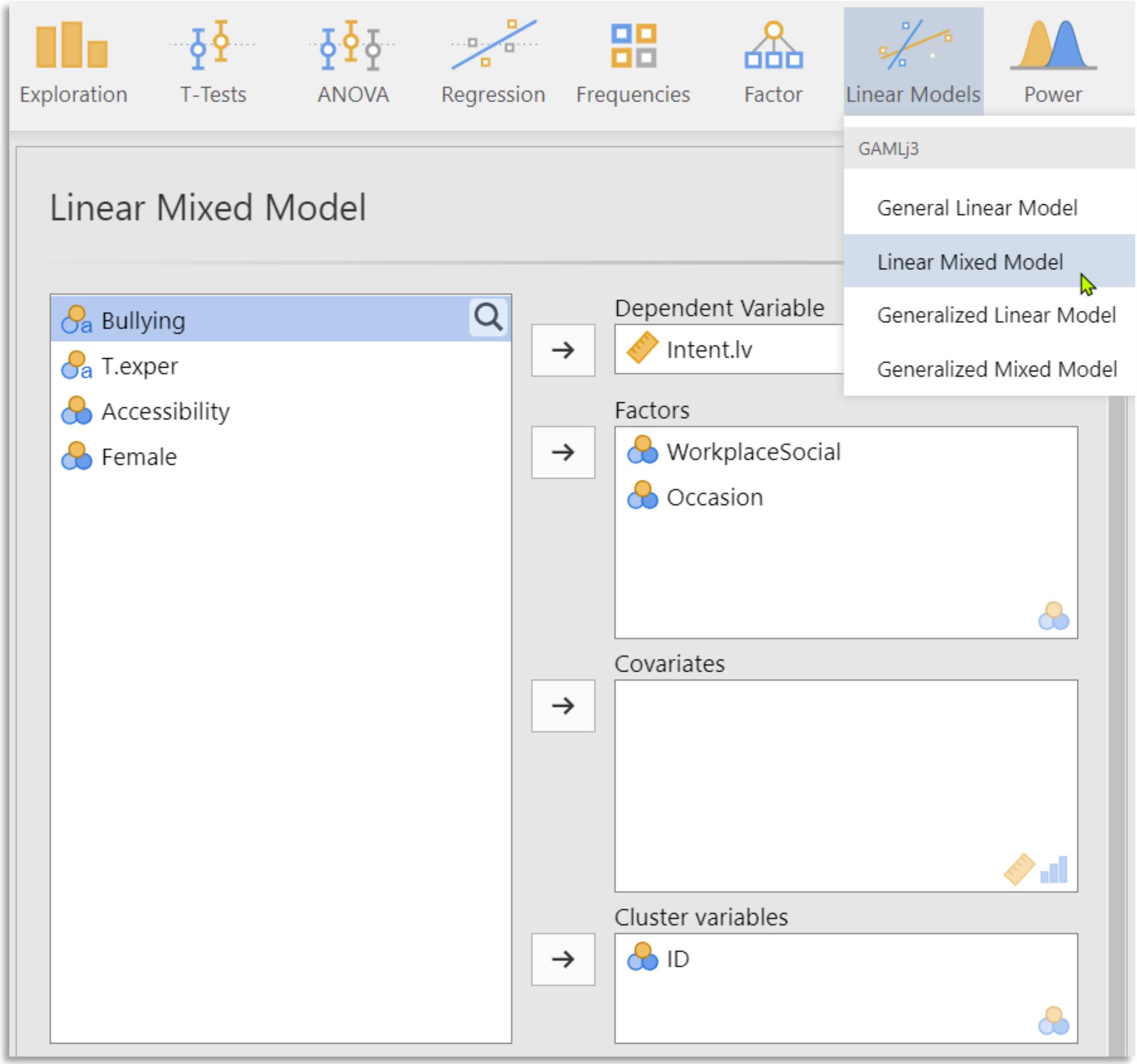
The between-group variable, Workplace-social event, is still a factor. Like ID, its levels are the same across all occasions for a person.

Linear Mixed Model: use **Long Format**

ID	Occasion	Intent.lv	WorkplaceSocial
1001	1	6.9	No
1001	2	16.6	No
1002	1	31.3	Yes
1002	2	28.7	Yes
1003	1	18.4	No
1003	2	18.2	No
1004	1	53.6	No
1004	2	88.0	No
1005	1	30.4	Yes
1005	2	26.8	Yes
1006	1	38.7	Yes

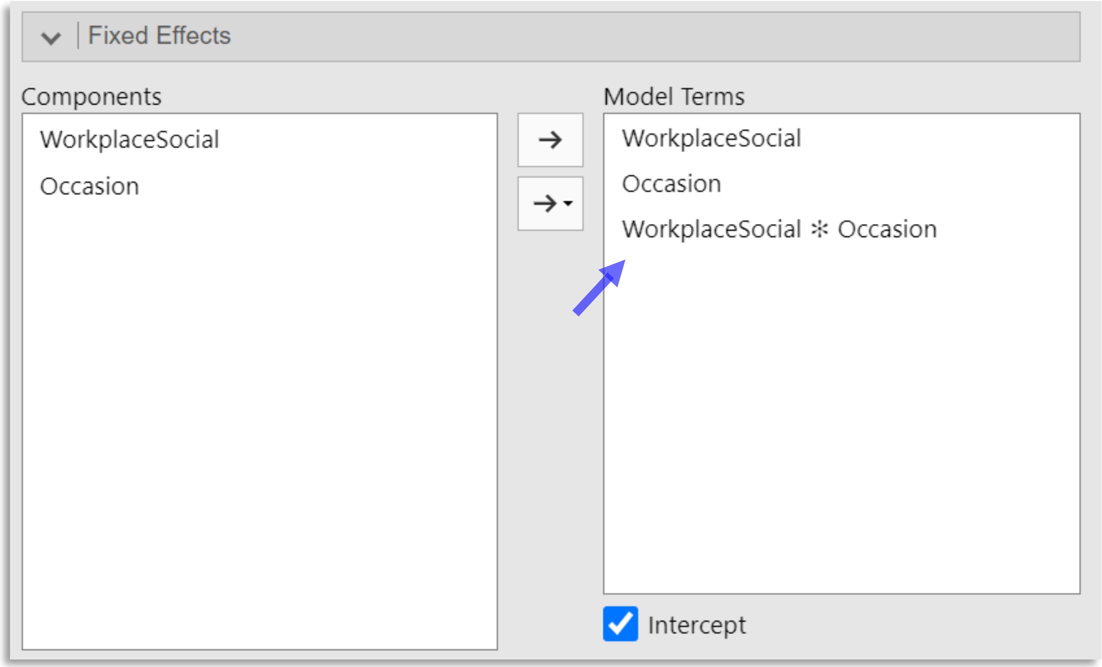
Linear Mixed Model Specification:

- Jamovi: Requires the GAMLj3 module. Linear Models → Linear Mixed Model
- Specify the variables
  - Both independent variables are categorical, so they are in the Factors field.
- **ID is a cluster variable.**
  - This is important because it tells Jamovi that each row that has the same ID is considered a cluster.
  - This is how the repeated measure, Occasion, gets modeled to be within each person.



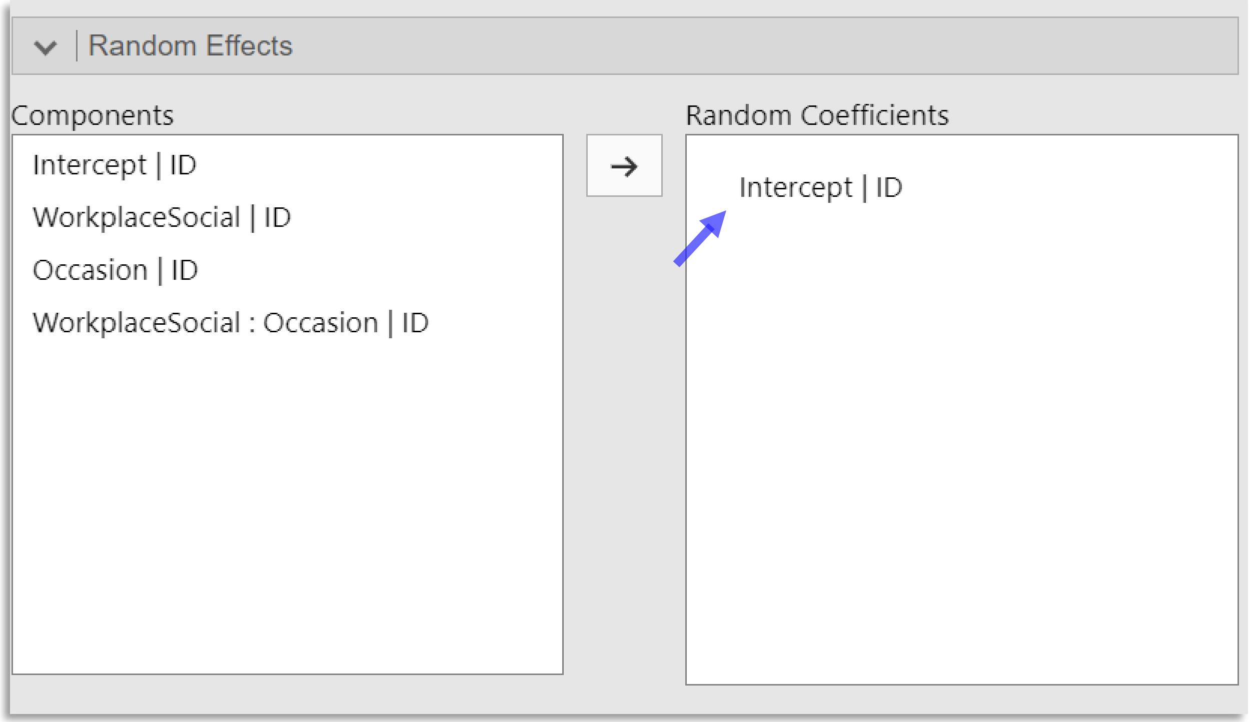
Specify the Fixed and Random Effects

Our hypothesis is about the interaction, so we should be sure it is included in the model.



H1: Teachers who participate in the workplace-social event series will become less intent to leave compared to their counterparts.

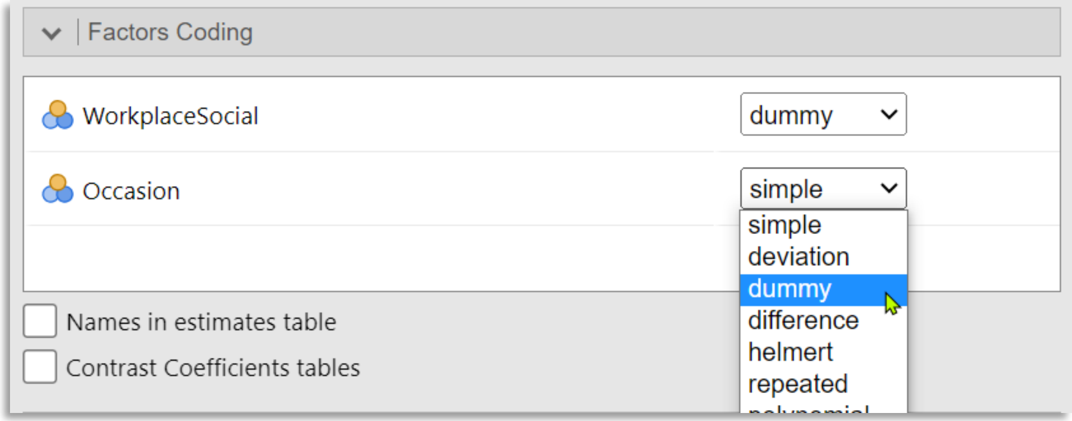
Because people individually differ in their baseline intent to leave, and because a person’s multiple observations are more similar to each other (correlated) than to other people’s observations, we need to include random intercepts in our model specification.

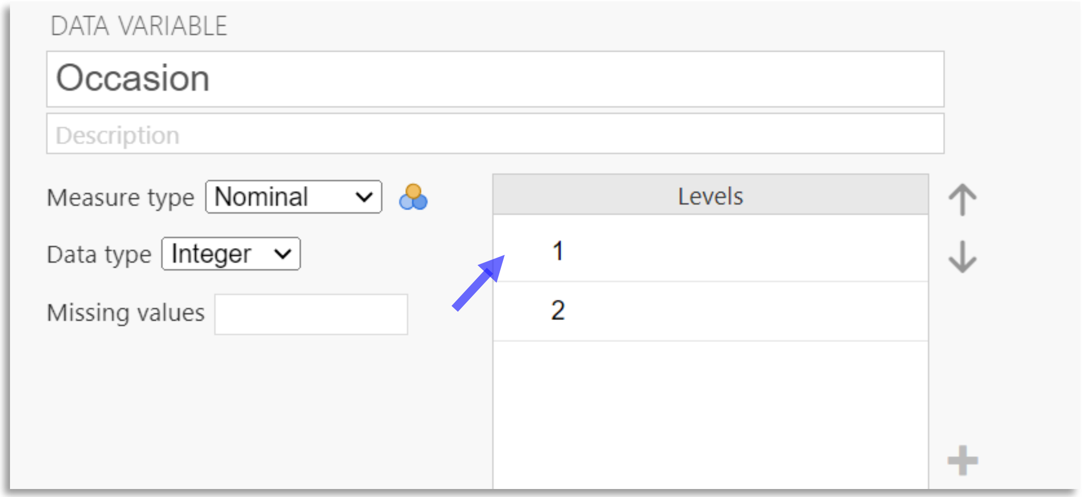


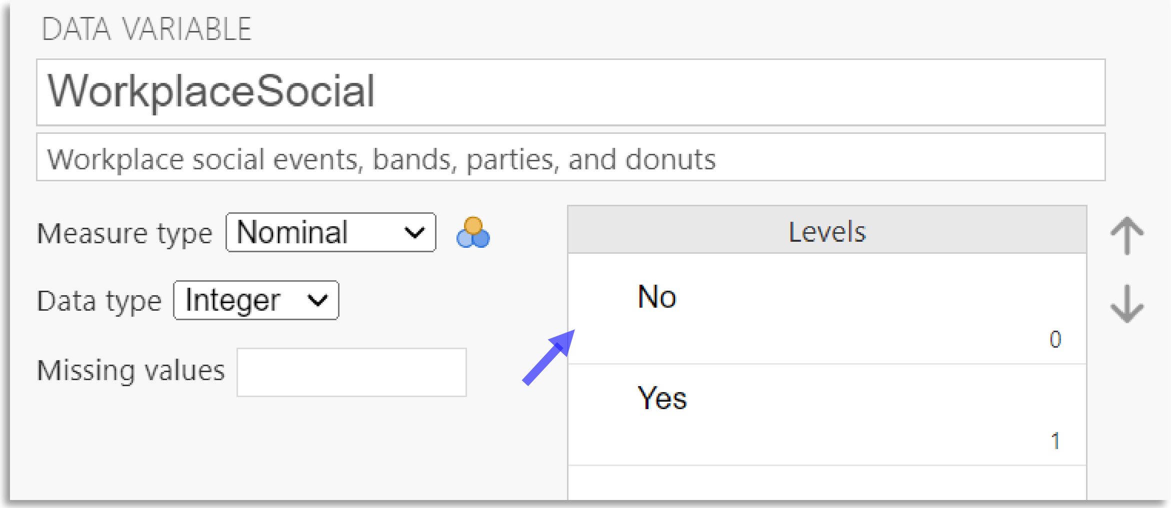
Specify the Type of Factor Coding

We need to check that the reference level is the top-most category in each factor.

For us, dummy coding is easy to interpret.

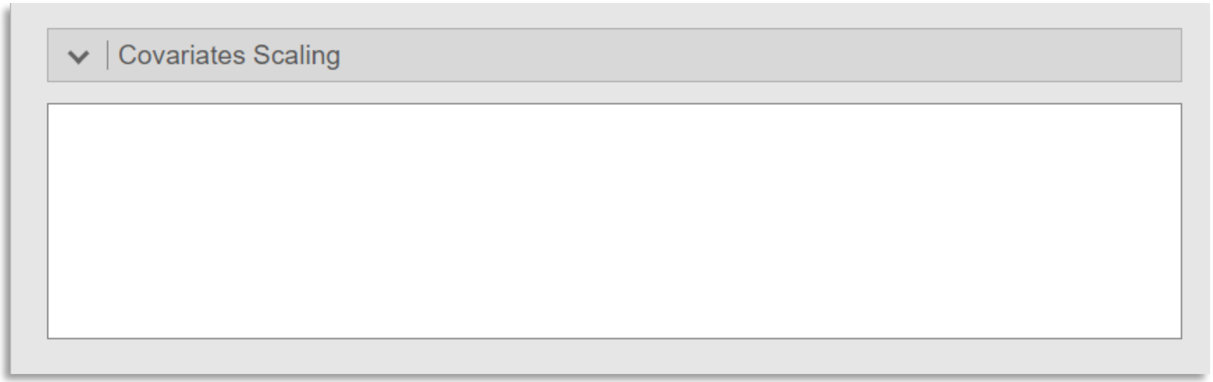






Specify the Covariate Scaling

- If we have any covariates, it helps our interpretation if they are on their original scale. We have no continuous independent variables in our model, so this is blank.



Examine Our Output for a Significant Interaction

- We see that Workplace × Occasion is statistically significant.
- The change in intent-to-leave depends on which group the teachers are in.

Fixed Effects Omnibus Tests

	F	df	df (res)	p
WorkplaceSocial	2.71	1	158.00	0.102
Occasion	24.87	1	158.00	< .001
WorkplaceSocial ✻ Occasion	24.75	1	158.00	< .001

Examine Our Output for a Significant Interaction

- Teachers in the workplace social condition showed a statistically significantly greater decrease in intent-to-leave over time, with an additional 16.37-point reduction compared to the control group.

H1: Teachers who participate in the workplace-social event series will become less intent to leave compared to their counterparts.

Parameter Estimates (Fixed coefficients)

				95% Confidence Intervals				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	44.66	3.16	38.45	50.88	205.77	14.14	< .001
WorkplaceSocial1	Yes - No	15.07	4.49	6.22	23.91	205.77	3.35	< .001
Occasion1	2 - 1	−0.02	2.31	−4.57	4.53	158.00	−0.01	0.993
WorkplaceSocial1 ✻ Occasion1	(Yes - No) ✻ (2 - 1)	−16.37	3.29	−22.84	−9.90	158.00	−4.97	< .001

- We have answered our research question with a single statistical test.

Linear Mixed Results Writeup

To examine whether teachers who participate in the workplace-social event series would become less intent to leave compared to their counterparts, we conducted a linear mixed model using the GAMLj3 module (Gallucci, 2019) in Jamovi (2024). Teachers in the workplace social condition showed a statistically significantly greater decrease in intent-to-leave over time, with an additional 16.37-point reduction compared to their control-group counterparts, $t(158) = -4.97$, $p < .001$.

References

Gallucci, M. (2019). GAMLj3: General analyses for linear models. (Version 3.6.0) [jamovi module]. https://gamlj.github.io/.

The jamovi project (2024). jamovi. (Version 2.6.44) [Computer Software]. https://www.jamovi.org.

Part 2: A New Application of the Linear Mixed Model

- Analysis in a Comparative Interrupted Time Series (CITS) design.
  - Multiple measures over time
  - Time points before and during or after the intervention
  - Treatment and control groups
  - Answers questions like

    "Did the intervention lead to an immediate change?"

    "Did the intervention change the rate of improvement?"

Treatment Group: O_1 O_2 O_3 O_4 O_5 X O_6 O_7 O_8 O_9 O_{10}

Control Group: O_1 O_2 O_3 O_4 O_5 O_6 O_7 O_8 O_9 O_{10}

Intervention time period (D): D = 0 D = 1

Example Scenario: A Reading Intervention for IEP students

The study design includes

The intervention: Each school day, from January to May, each IEP student is provided with a 20-minute targeted reading intervention.

Outcome: Words Correct Per Minute (WCPM), as a measure of reading fluency once every month, from October through May.

Treatment group: Students with IEPs, who receive the intervention from Jan to May

Control group: Students without IEPs, who do not receive the intervention at any time

Research question: Is a targeted daily reading intervention effective for students with IEPs, and does it help reduce the reading gap between IEP and non-IEP students?

- Do students receiving the treatment get an immediate boost, compared to those not receiving the treatment?
- Do students in the intervention show more growth in improvement, compared to their counterparts?

Reading Intervention Timeline

Data Collection Schedule:

Month	Time	Implementation (D)	Intervention Status
October	-2	0	Pre-intervention
November	-1	0	Pre-intervention
January	0	1	Intervention begins
February	+1	1	Intervention continues
March	+2	1	Intervention continues
April	+3	1	Intervention continues
May	+4	1	Intervention continues

Model Specification Equation

$$
\begin{aligned}
Y_{ij} = b_0 \quad &+ \quad b_1(T) + b_2(G) + b_3(D) \\
&+ \quad b_4(T \times G) + b_5(T \times D) \\
&+ \quad b_6(G \times D) \\
&+ \quad b_7(T \times G \times D) \\
&+ \quad u_{0i} + \varepsilon_{ij}
\end{aligned}
$$

Where

$$
\begin{aligned}
Y_{ij} &= \text{WCPM for student } i \text{ at time } j \\
T &= \text{Time, months from intervention start} \\
G &= \text{Group (Intervention Group)} \\
D &= \text{Implementation time} \\
u_{0i} &= \text{Random intercept} \\
\varepsilon_{ij} &= \text{Error term}
\end{aligned}
$$

Model Specification Equation, further explained

$$
\begin{aligned}
Y_{ij} = b_0 \quad &+ \quad b_1(T) + b_2(G) + b_3(D) \\
&+ \quad b_4(T \times G) + b_5(T \times D) \\
&+ \quad b_6(G \times D) \\
&+ \quad b_7(T \times G \times D) \\
&+ \quad u_{0i} + \varepsilon_{ij}
\end{aligned}
$$

Where:

$$
\begin{aligned}
Y_{ij} &= \text{WCPM score for student } i \text{ at time } j \\
b_0 &= \text{Intercept (baseline WCPM for non-IEP students at Time 0)} \\
b_1 &= \text{Linear time trend} \\
b_2 &= \text{Group effect (IEP vs. non-IEP difference)} \\
b_3 &= \text{Level change at intervention start} \\
b_4 &= \text{Group difference in time trend} \\
b_5 &= \text{Slope change after intervention} \\
b_6 &= \text{Group difference in level change} \\
b_7 &= \text{Group difference in slope change} \\
T &= \text{Time (months relative to intervention start)} \\
G &= \text{Group (0 = non-IEP, 1 = IEP)} \\
D &= \text{Implementation time (0 = pre, 1 = post)} \\
u_{0i} &= \text{Random intercept for student } i \\
\varepsilon_{ij} &= \text{Residual error for student } i \text{ at time } j
\end{aligned}
$$

Data Set

Categorical variables:

Student ID (cluster variable)

Group

D

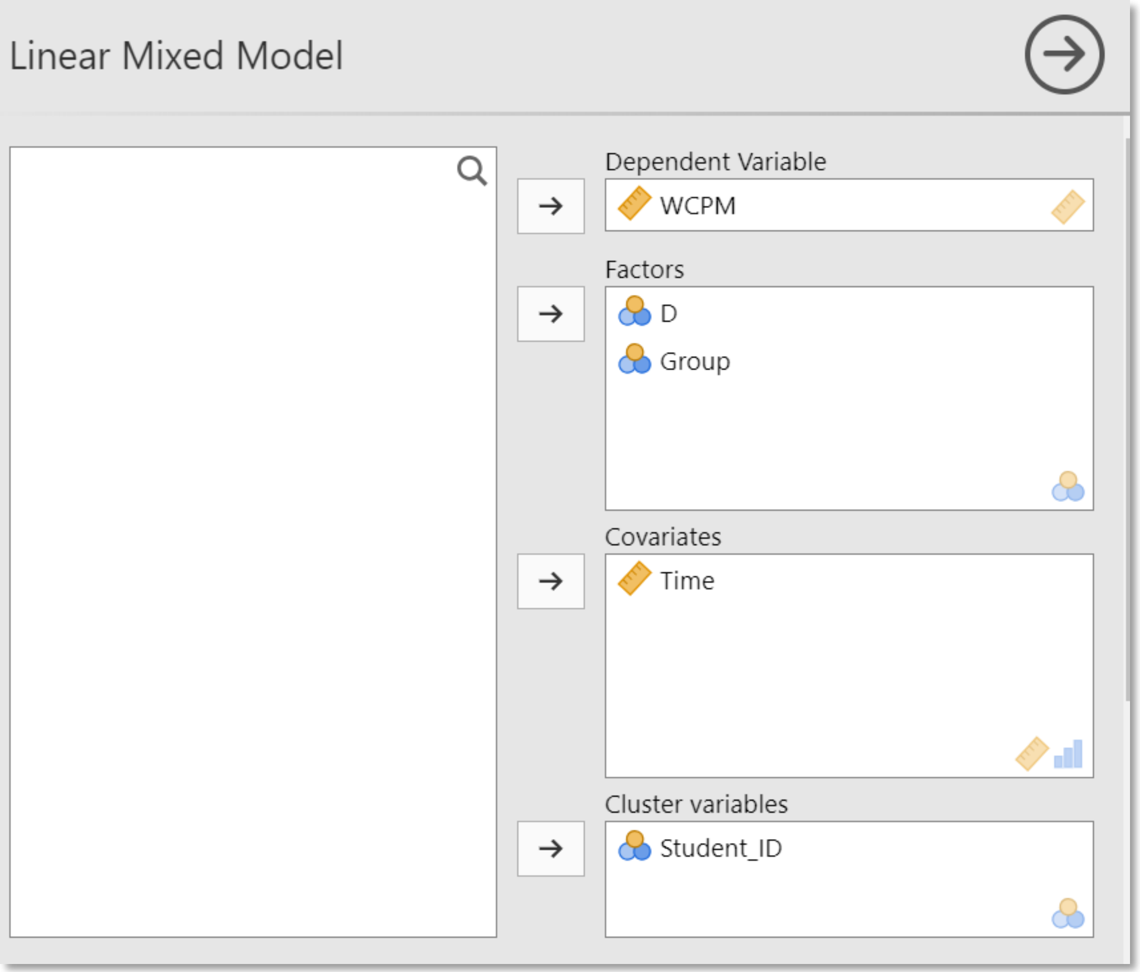
Continuous variables:

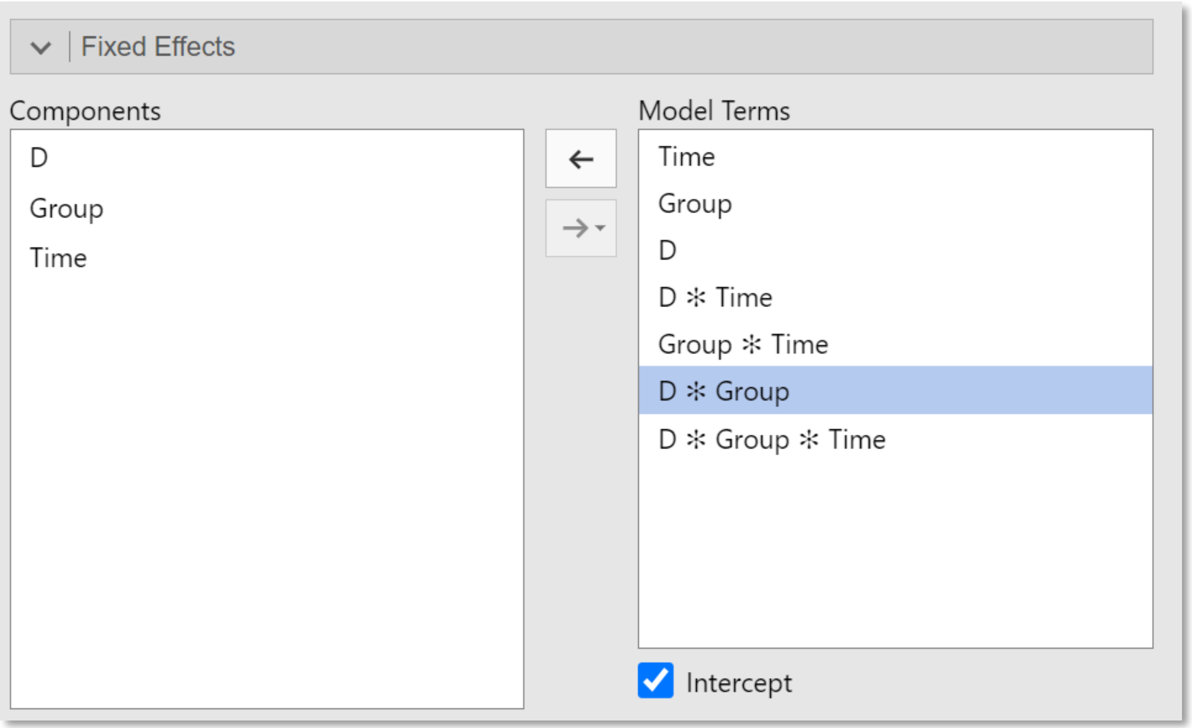
Time

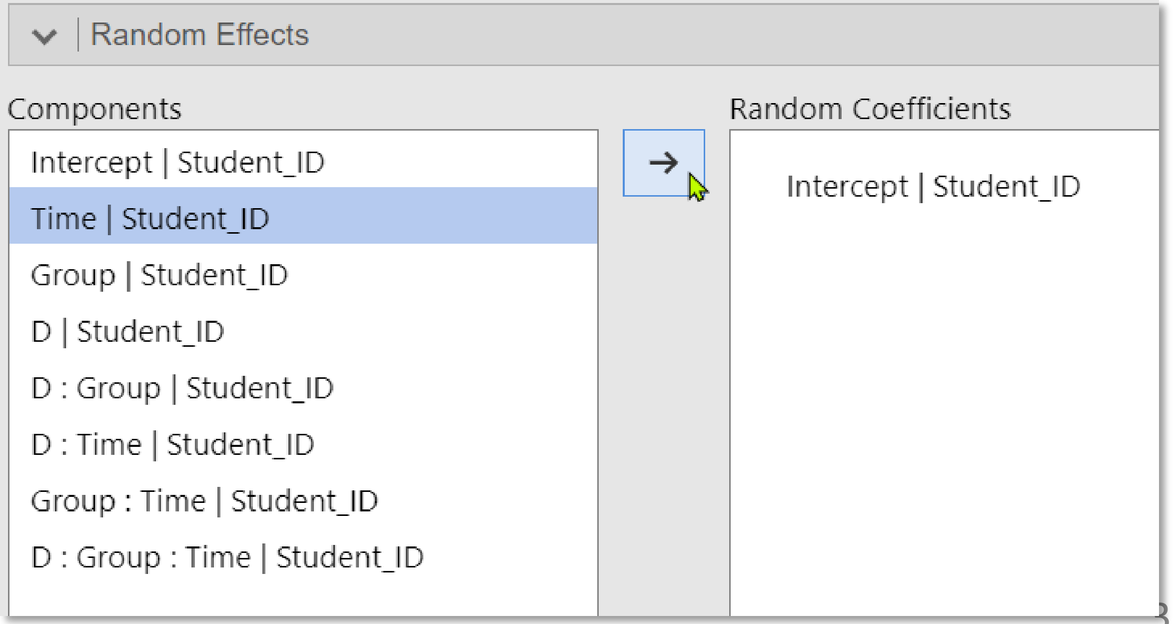
WCPM

	Student_ID	Group	Time	ImplemTime(D)	WCPM
1	1	IEP	-2	No	26.80
2	1	IEP	-1	No	18.79
3	1	IEP	0	Yes	32.99
4	1	IEP	1	Yes	38.71
5	1	IEP	2	Yes	53.13
6	1	IEP	3	Yes	58.28
7	1	IEP	4	Yes	64.33
8	2	IEP	-2	No	28.94
9	2	IEP	-1	No	30.31
10	2	IEP	0	Yes	38.56
11	2	IEP	1	Yes	57.32
12	2	IEP	2	Yes	57.15
13	2	IEP	3	Yes	67.40
14	2	IEP	4	Yes	73.12
15	3	No IEP	-2	No	67.21

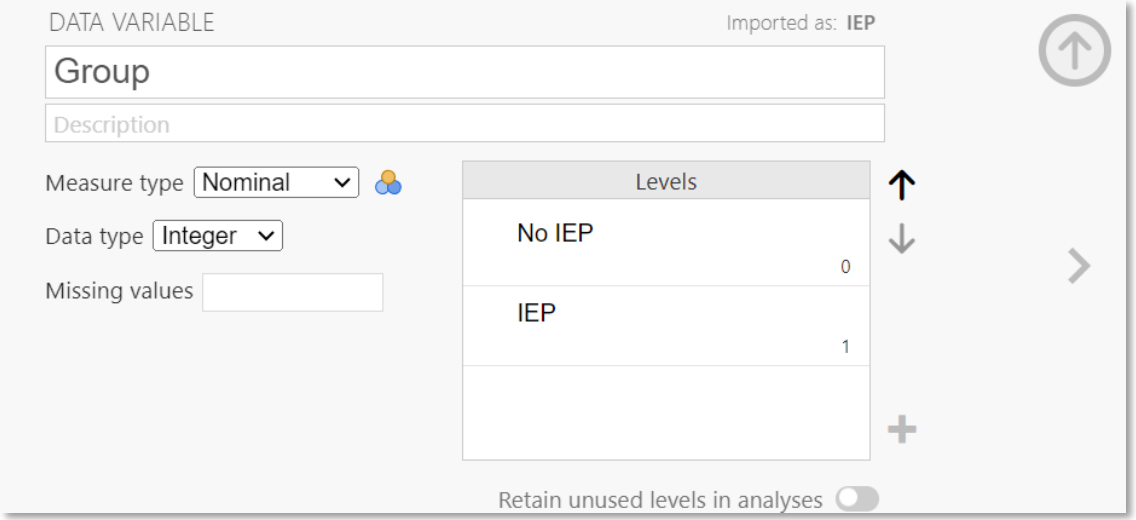
Specification in Jamovi's GAMLj3 Linear Mixed Procedures

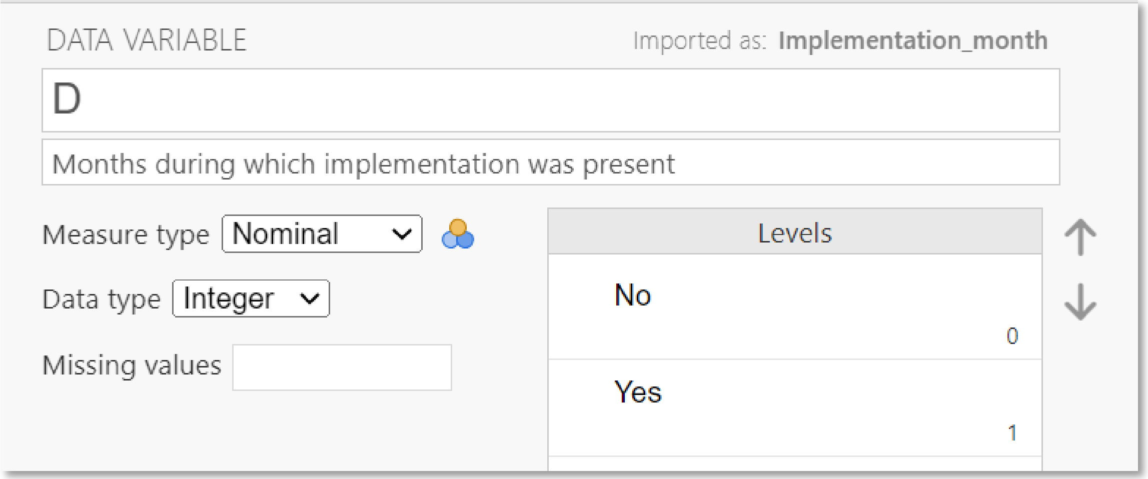






Specification in Jamovi's GAMLj3 Linear Mixed Procedures





Specification in Jamovi's GAMLj3 Linear Mixed Procedures

Output

Fixed Effects Omnibus Tests

	F	df	df (res)	p
Time	60.78	1	714.00	< .001
Group	89.54	1	167.07	< .001
D	14.28	1	714.00	< .001
Time ✻ D	47.50	1	714.00	< .001
Time ✻ Group	2.51	1	714.00	0.114
Group ✻ D	2.25	1	714.00	0.134
Time ✻ Group ✻ D	4.52	1	714.00	0.034

Output

Parameter Estimates (Fixed coefficients)

Names	Effect	Estimate	SE	95% Confidence Intervals Lower	95% Confidence Intervals Upper	df	t	p
(Intercept)	(Intercept)	63.22	2.00	59.29	67.15	330.01	31.61	< .001
Time	Time	0.61	0.84	−1.04	2.26	714.00	0.73	0.467
Group1	IEP - No IEP	−32.84	3.94	−40.57	−25.12	330.01	−8.35	< .001
D1	Yes - No	3.15	1.40	0.39	5.91	714.00	2.24	0.025
Time ✻ D1	Time ✻ (Yes - No)	4.03	0.86	2.35	5.72	714.00	4.69	< .001
Time ✻ Group1	Time ✻ (IEP - No IEP)	−0.46	1.65	−3.70	2.79	714.00	−0.28	0.782
Group1 ✻ D1	(IEP - No IEP) ✻ (Yes - No)	4.14	2.76	−1.28	9.57	714.00	1.50	0.134
Time ✻ Group1 ✻ D1	Time ✻ (IEP - No IEP) ✻ (Yes - No)	3.60	1.69	0.27	6.92	714.00	2.12	0.034

[5]

Key Statistical Tests for Addressing the Research Questions

Research question: Is a targeted daily reading intervention effective for students with IEPs, and does it help reduce the reading gap between IEP and non-IEP students?

- Do students receiving the treatment get an immediate boost, compared to those not receiving the treatment?
  - **b6(Group × D):** No significant immediate effect difference, estimate $= 4.14$ words per minute, $p = .134$.
- Do students in the intervention show more growth in improvement, compared to their counterparts?
  - **b7(Time × Group × D):** Positive significant slope difference, estimate = 3.60 words per minute, $p = .034$.

Linear Mixed Results Writeup

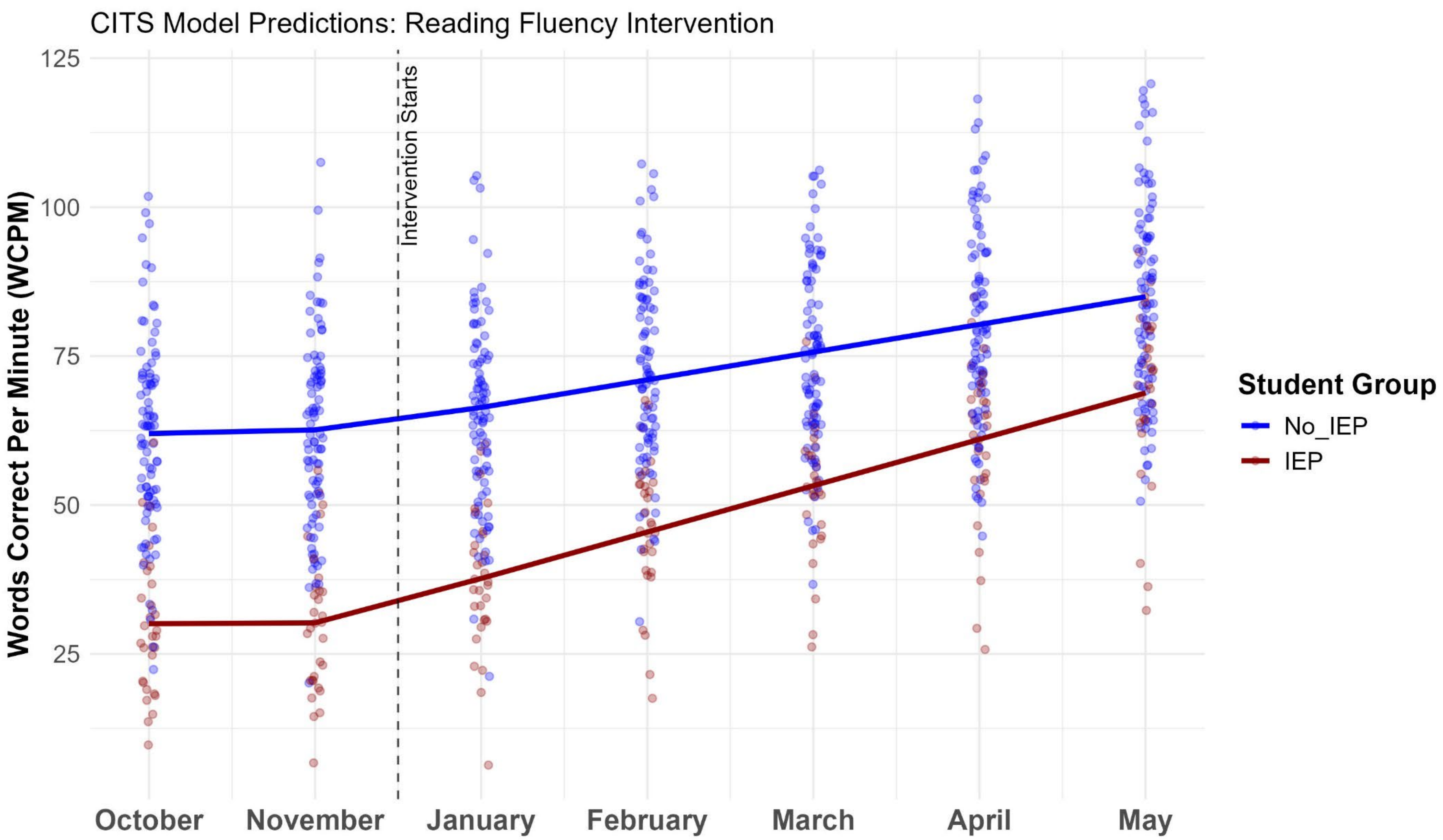
To examine whether a daily reading intervention improves reading fluency for students with IEPs, we conducted a comparative interrupted time series analysis using a linear mixed model in the GAMLj3 module (Gallucci, 2019) in Jamovi (2024). Students with IEPs did not show a significantly different immediate improvement in reading fluency when the intervention began compared to non-IEP students, $t(714) = 1.50$, $p = .134$. However, IEP students experienced a statistically significantly greater rate of monthly improvement after the intervention started, gaining an additional 3.60 words per minute each month compared to non-IEP students, $t(714) = 2.12$, $p = .034$.

References

Gallucci, M. (2019). GAMLj3: General analyses for linear models. (Version 3.6.0) [jamovi module]. https://gamlj.github.io/.

The jamovi project (2024). jamovi. (Version 2.6.44) [Computer Software]. https://www.jamovi.org.

We could plot the predicted equation, such as in Excel (or R)



Key Takeaways

- Linear mixed models take regression to the next level.
- Mixed ANOVA = a special case of a linear mixed model
- When to Choose Linear Mixed Models
  - When we have multiple measures over time, along with covariates that also vary over time
  - When we have nested data, such as students nested within classes
- CITS Analysis: A quasi-experimental approach for examining program effectiveness
  - Can compare groups in their change over time when an intervention is introduced part way through.
  - Can addresses equity questions about whether a program serves to close the gaps