

The Comparison Between Repeated Measures Anova And Multilevel Modeling To Investigate The Effectiveness Of Language Instruction

Eunjeong Park

Sunchon National University, Sunchon, South Korea

ABSTRACT

The purpose of the current study is to compare two different statistical analyses—repeated measures analysis of variance (ANOVA) and multilevel modeling (MLM)—with regard to international second language (L2) students' lexico-grammatical writing gains in the interventions. Fifty college students' lexico-grammatical writing gains in the pre-test, immediate and delayed post-tests were collected from three groups—one control and two intervention groups. Research questions include: (1) Is there a change in the dependent variable over time through repeated measures ANOVA?; (2) How much on average do the individuals change over time through MLM?; and (3) What are the differences of the results between repeated measures ANOVA and MLM? Repeated measures ANOVA and MLM were conducted separately to answer the research questions. Both analyses revealed that there was no difference of the interventions over time. In terms of methodological considerations, however, the two analyses exhibited the results in a different manner. Repeated measures ANOVA requires several assumptions (e.g., homoschedasticity and sphericity) while MLM does not need the requirement of the data as long as it contains different periodic dataset. Furthermore, MLM enables researchers to investigate the rate of individual growth. This paper offers the methodological insight of adequate quantitative analyses to investigate the effectiveness of language acquisition and learning.

Keywords: Repeated measures ANOVA, multilevel modeling, language instruction.

INTRODUCTION

Interest in multiple levels of theory and analysis has been on the increase in language research. This increase in multilevel research has led to a variety of uses of analytical techniques, such as repeated measures analyses (e.g., repeated measures ANOVA) and multilevel modeling (MLM). Language researchers may encounter challenges of estimating repeated measures with different analysis tools. Many researchers who have a background in regression models come to multilevel modeling (MLM), while others utilize ANOVA methods for analyzing experiments with random effects in experimental disciplines [1]. Any statistical technique would have two sides of the coin. In particular, repeated measures ANOVA and MLM are popular in the research fields. The repeated measures ANOVA design allows statistical inference with fewer number of subjects and offers efficiency in terms of a quick analysis and less cost of training. MLM allows researchers to study

both intra- and inter-individual differences in the growth parameters (e.g., slopes and intercepts). It is also flexible in handling missing data. Given the complexity of analyzing longitudinal data, the current study aimed to compare the results of repeated measures ANOVA and MLM and examine the feasibility of both analyses for the effects of second language instruction.

Intentional and incidental language learning

For the pedagogical development, this study employed intentional and incidental language instruction to second language writing classes. Different types of language learning can remain on a continuum between incidental and intentional [2] since attention itself is not a dichotomous construct [3]. Therefore, both intentional and incidental learning can occur with some degree of attention [4]. Hulstijn [5,6] also claims that intentional or incidental learning involves some amount of attention. Unlike the argument about the language instruction on a continuum, intentional and incidental learning are regarded as “complementary activities, each one enhancing the learning that comes from the other” (p. 232) [7]. Nation and Webb [8] justified the reason for distinguishing intentional and incidental learning: the investigation of the amount and efficiency of learning. This study operationalized intervention terms based on intentional and incidental language learning, in order to explore the effectiveness of lexical bundle instruction.

Lexical bundles

Second language researchers have investigated lexicogrammar with diverse components of a language. One of the features within lexicogrammar is a “lexical bundle” [9]. Lexical bundles are the most frequent sequences of multiple words in a register indicating formulaicity of lexicogrammar in a language. Because of the nature of ubiquity in language, it seems natural and easy to acquire lexical bundles. However, producing lexical bundles in both spoken and written registers is challenging for second language students in diverse disciplines and levels of expertise [10-12]. Lexical bundles can be a marker of proficiency in a second language [10, 11]. Many language researchers [13-16] also argue that becoming adapted to the rhetorical styles and writing expectations of different disciplines is important in gaining membership in the academic community. Therefore, teaching lexical bundles would help second language learners improve the use of lexical bundles, which eventually would enhance their writing proficiency level and encourage them to be identified as an eligible member in academia.

Repeated measures analysis of variance (ANOVA)

Repeated measures Analysis of variance (ANOVA) is used to compare the final measures or the change scores between the baseline and final measurement between the treatment groups. It can be useful when researchers are interested in measuring trends over time.

There are several benefits of repeated measures ANOVA. First, it can have more statistical power, controlling for factors that cause variability between subjects. Second, repeated measures ANOVA can use fewer subjects to detect a desired effect size thanks to the greater statistical power. This also leads to cost-effectiveness in that fewer subjects need to be recruited or compensated to complete an entire process of the experiment. Lastly, it can track an effect over time—the learning curve for a task. Repeated measures ANOVA offers a complete description of the time effect due to the inclusion of all time points, allowing statistical assessment, such as treatment $\times$ time interaction effect [17].

Despite these benefits, repeated measures ANOVA has some challenges. This analysis requires full data on everyone at each time point. Therefore, missing data on the final measure becomes a serious problem. The missingness of the final measure may cause a lack of representativeness of the population and sample bias [17]. Thus, end-point analysis can lead to misleading results from varying dropout rates between the treatment groups. It also requires the assumption of sphericity that equal variability of the measurement at each time point is expected. This assumption is vulnerable because consecutive observations on the same subject are likely to be highly correlated. This causes the inflation of the Type I error rate and the statistical significance of the treatment $\times$ time effect can be overestimated. To resolve this issue, the Greenhouse-Geisser and Huynh-Feldt corrections to the significance tests are employed to repeated measures ANOVA. These corrections diminish the numerator degrees of freedom to adjust p-value upward and make the Type I error rate closer to the alpha level [17].

Multilevel modeling (MLM)

Multilevel modeling (MLM) is commonly used for repeated measures designs because it presents an alternative approach to analyzing repeated data with four main advantages. First, MLM has less stringent assumptions. MLM can be used if the assumptions of homoscedasticity are violated for repeated measures ANOVA. Therefore, the assumptions of homogeneity of variance are not necessary in MLM, which tends to yield more power. Second, MLM can be used for higher-order sampling procedures by examining repeated measures within subjects or the third level of analysis. Third, missing data is manageable in MLM without causing additional complications. It is not needed to exclude the data if they are missing a single data point. Last, there is a flexibility of the variation in the timing of data collection (i.e., variable timing vs. fixed timing). Depending on participant availability and other scheduling issues, variation of data collection may occur. In other words, measurement periods can vary among the individuals. This variation may be addressed in MLM by adding “age” into the regression equation. Furthermore, MLM does not require equal intervals between measurement points. MLM also resolves the issue of measurement error in pre-test or post-test scores that reduces the precision of the estimate of the treatment effect [18, 19].

METHOD

The purpose of the current study is to demonstrate and compare the results of longitudinal data of second language instruction via both repeated measures ANOVA and MLM.

Research Questions

- 1) Is there a change in the dependent variable over time through repeated measures ANOVA?
- 2) How much on average do the individuals change over time through MLM?
- 3) What are the differences of the results between repeated measures ANOVA and MLM?

Participants

Convenience and purposive samplings were employed for the participant recruitment in this study. Purposive sampling, in particular, includes several characteristics: (1) a selection of participants based on specific purposes related to research questions, (2) expert judgment of researchers and informants, and (3) a small sample size (usually 30 or fewer cases) [20]. Recruiting the research participants was available by asking the ESL Program Office at the Midwestern university. Three ESL classes were assigned according to the availability and the research criteria (i.e., the population of international second language students) as purposive sampling. The sampling was made during

the spring semester in 2017 with fifty college students taking the three intact ESL composition classes. The participants were international undergraduate students who did not permanently reside in the United States but have a VISA for study abroad. The homogeneity among the groups was ensured according to gender, age, ethnicity, the academic record (i.e., GPA). The students who did not agree on the consent were not included in the study.

Fifty students were assigned to one control and two intervention groups. There were 22 male students (44 %) and 28 female students (56 %). 44 students (88 %) were from China, while six (12 %) of them were from Malaysia. The range of the age was 18 to 24 years old ($M=19$, $SD=1.40$). Their field of study varied as shown in Table 1.

Table 1. Study participants in each group

	Intentional Group	Incidental Group	Control Group
Total number	19	18	13
Mean age	20 ($SD=1.7$)	19 ($SD=0.9$)	19 ($SD=0.9$)
Gender			
Male	11	7	4
Female	8	11	9
L1			
Chinese	16	15	13
Malaysian	3	3	0
Field of study			
Business	4	7	5
Sciences	6	4	2
Mathematics		5	2
Engineering	3		2
Sport industry	1	1	1
Design	2		
Others	3	1	1

Variables

Test Scores of Lexico-grammatical Writing Gains

Variables used in this paper were pre-, immediate post-, and delayed post-tests of lexico-grammatical writing gains from the interventions (i.e., intentional and incidental) as the outcome variable. The test included 20 cloze test items; thus, the scores were measured on the ratio scale.

Interventions of Intentional and Incidental Lexical Bundle Learning

This study operationalized terms of the interventions: intentional learning intervention, incidental learning intervention, and a control group. The purposes of intentional instruction included acquiring lexical bundles with explicit definitions and examples of lexical bundles and analyzing lexical bundles. The intentional learning intervention had a clear intent with the observable evidence (e.g., explicit explanation of lexical bundles). The incidental learning intervention was based on three general processes of noticing, retrieval, and generation [7]. The purposes of incidental instruction included noticing lexical bundles implicitly in the process of instruction and getting familiarized with them. A control group was designated in a regular ESL writing classroom. This control group as a reference set was compared to the two intervention groups.

Each of the intentional and incidental instruction included five 20-minute lexical bundle activity sessions per week during the whole semester (spring in 2017), approximately over a period of five months. The interventions were carried out for 12 weeks: a pre-test period (1 week), interventions (5 weeks), an immediate post-test period (1 week), and a delayed post-test period (after 5 weeks) in Table 2. The control group had curriculum-based lessons in the composition class.

Table 2. The procedure of the interventions

	Pre-test	Interventions (Intentional/incidental)	Immediate post-test	Delayed post-test
Duration	1 week before the interventions	5 weeks 5 days × 20 min. a session (1 session/week)	1 week after the interventions	In 5 weeks after the post-test

ANALYSIS

Repeated measures ANOVA

For repeated measures ANOVA, General Linear Model was selected from IBM SPSS Statistics 24 (i.e., Statistical Package for the Social Science) software program. SPSS Statistics makes it easy to test whether data has met or failed the assumption of sphericity—the variances of all paired differences to be equal. Mauchly's test of sphericity is performed in SPSS Statistics. Pairwise comparisons indicate the processes of comparing entities in pairs to assess which has a greater amount of quantitative property.

The repeated measures ANOVA tests for any differences between related population means. The null hypothesis (H_0) states that the means are equal:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

The alternative hypothesis (H_A) tests if the related population means are not equal (at least one mean is different to another mean):

H_A : at least two means are significantly different

For this paper, the null hypothesis (H_0) is that scores of the lexical bundle test (i.e., lexico-grammatical writing gains) are the same at all time points (pre-, immediate post-, and delayed post-test). The alternative hypothesis is that the mean scores are significantly different at one or more time points. Since repeated measures ANOVA only checks the differences of the outcome variable for the trend over time, it will not inform where the differences between groups lie as it is an omnibus statistical test. Repeated measures ANOVA can further partition the error term, reducing its size. Therefore, it increases the value of the F-statistic leading to an increase in the power of the test to detect significant differences between means. As the same subjects in each group are used, the variability of between-subjects is removed.

Multilevel modeling

MLM was conducted by HLM (i.e., Hierarchical Linear Modeling) software program that enables researchers to generate a linear model with explanatory variables accounting for variations at each level. HLM estimates model coefficients and predicts the random effects at each level. Due to the capacity of analyzing a hierarchical structure, HLM can be applied to a longitudinal analysis, in

which an individual's repeated measurements can be nested within the individuals being studied. HLM allows for continuous, count, ordinal, and nominal outcome variables, assuming a functional relationship between the expectation of the outcome and a linear combination of a set of explanatory variables by a suitable link function: the identity link (continuous outcomes) or logit link (binary outcomes) [21].

In an HLM individual growth model, level 1 describes an individual's growth trajectory across time. A simple two-level linear growth model is illustrated below.

Simple Model

Level 1: $LBTEST_{ij} = \beta_{0j} + \beta_{1j} * (TIME_{ij}) + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} * (INTERVEN1_j) + \gamma_{02} * (INTERVEN2_j) + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} * (INTERVEN1_j) + \gamma_{12} * (INTERVEN2_j) + u_{1j}$

INTERVEN1: Intentional Group

INTERVEN2: Incidental Group

Mixed Model

$LBTEST_{ij} = \gamma_{00} + \gamma_{01} * INTERVEN1_j + \gamma_{02} * INTERVEN2_j + \gamma_{10} * TIME_{ij} + \gamma_{11} * INTERVEN1_j * TIME_{ij} + \gamma_{12} * INTERVEN2_j * TIME_{ij} + u_{0j} + u_{1j} * TIME_{ij} + r_{ij}$

The observations across time are nested within persons. The explanation of the basic two-level HLM model is based on *Handbook of Quantitative Methods for Educational Research* [23]. The level 1 equation represents individual trajectories within individual variability across time. The dependent variable is the lexico-grammatical writing gains for the individuals. We predict three things: 1) the intercept (β_{0j}) when time is zero; 2) the product of a constant rate of change and time, and 3) the individual error (r_{ij}). The time slope (β_{1j}) represents the linear rate of change over time. The deviation (r_{ij}) of an individual from the predicted trajectory can be regarded as the measurement error related to the individual's estimate at the time point. The level 2 equation shows the average growth trajectories across individuals and deviations from the averages. The randomly varying intercept (β_{0j}) is predicted by an overall intercept (γ_{00}), the effects of level 2 variables on the intercept, and the level 2 residual (u_{0j}), which indicates the difference between the model implied intercept and the individual's observed intercept. The randomly varying slope (β_{1j}) is predicted by an overall intercept (γ_{10}), the effects of level 2 variables on the linear growth slope, and the level 2 residual (u_{1j}), which indicates the difference between the person's model predicted linear growth slope and the individual's growth slope. The level 2 model explains variation in individuals' growth trajectories with person level predictors, such as gender and treatment groups.

RESULTS

Descriptive statistics of the data

Descriptive statistics are presented for the fifty students' pre-test and two post-tests (Table 3). The mean score of the pre-tests ranged from 4 to 18 ($M=11.86$, $SD=3.75$). The mean score of the post-tests ranged from 3 to 20 ($M=14.38$, $SD=3.76$). The mean score of the delayed post-tests ranged from 6 to 20 ($M=14.88$, $SD=3.47$).

Table 3. Descriptive statistics

	Pre-test					Post-test				Delayed post-test			
	N	M	SD	Min.	Max.	M	SD	Min.	Max.	M	SD	Min.	Max.
Intentional	19	12.21	4.26	5	18	14.47	3.44	8	20	16.21	3.36	6	20
Incidental	18	12.83	3.10	4	17	16.17	3.10	10	20	14.89	3.55	6	20
Control	13	10.00	3.34	5	18	11.77	3.77	3	16	12.92	2.75	7	17
Total	50	11.86	3.75	4	18	14.38	3.76	3	20	14.88	3.47	6	20

Note: M=Mean, SD=Standard Deviation

In order to examine the effect of the interventions, it is important to meet the assumptions of the univariate analysis, such as normality and homogeneity of variance. The normality of pre-test and post-test variables were first checked. The skewness and kurtosis statistics of variable residuals were all within ± 2 close to zero. Q-Q plots and histograms showed that the data were normally distributed as shown in Figure 1.

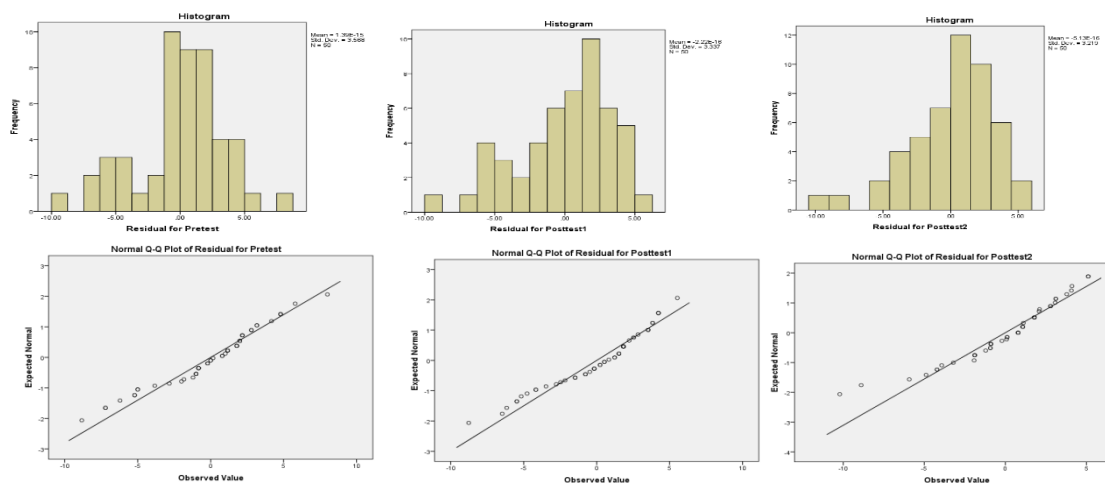


Figure 1. Histograms and Q-Q plots of residuals of pre-test and post-tests

Results from repeated measures ANOVA

The results of the test of sphericity indicate the assumption of sphericity was not violated. Based on the results of the within-subjects ANOVA ($F(2, 94)=18.49, p<.001$), we reject the null hypothesis. There was a significant difference in the lexical bundle writing test scores over time with a relatively large effect size (partial $\eta^2=.28$). There were statistically significant differences in the mean scores for the pre-test vs. the immediate post-test and the pre-test vs. the delayed post-test with the Bonferroni multiple comparison procedure. The between-subjects ANOVA showed a significant difference in the lexico-grammatical writing gains obtained by the participants receiving different interventions with a relatively large effect size ($F(2, 47)=5.49, p=.007$, partial $\eta^2=.189$) as shown in Table 4. Post-hoc tests of the Bonferroni multiple comparison indicate statistically significant differences in INTERVENTION for CONTROL vs. INTENTIONAL and CONTROL vs. INCIDENTAL groups with the Bonferroni multiple comparison procedure in Table 5. However, the interaction between TIME and INTERVENTION was not significant, indicating there was not a difference by INTERVENTION in lexico-grammatical writing gains of the participants.

Table 4. Tests of within- and between-subjects effects

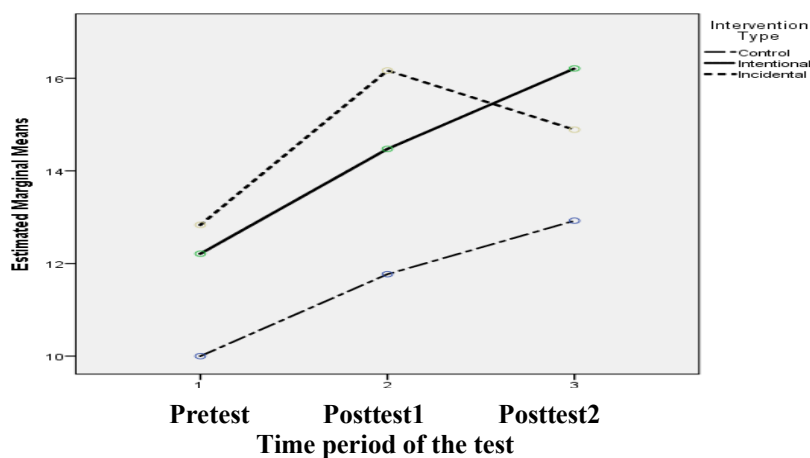
Measure: Lexico-grammatical Writing Gain Scores								
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent Parameter	Observed Power ^a
Within subjects:								
Lexico-grammatical Test	247.55	2	123.77	18.49	.000	.28	36.99	1.00
Time*Intervention	49.04	4	12.26	1.83	.129	.07	7.33	.54
Error (LB)	628.95	94	6.69					
Between subjects:								
Intervention	244.98	2	122.49	5.49	.007	.189	10.986	.83
Error	1048.11	47	22.30					

a. Computed using alpha = .05

Table 5. Post hoc tests: Multiple comparisons

Measure: Lexico-grammatical Writing Gain Scores						
Bonferroni						
(I) Intervention Type	(J) Intervention Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Control	Intentional	-2.73*	.981	.023	-5.17	-.30
	Incidental	-3.07*	.992	.010	-5.53	-.60

Based on observed means.
The error term is Mean Square(Error) = 7.433.
*. The mean difference is significant at the .05 level.

**Figure 2. Average raw lexico-grammatical writing gain scores by the intervention type**

Results from HLM

The level 2 model of the longitudinal data shows how the intercepts and slopes are moderated by what INTERVENTIONS the students had. Restricted maximum likelihood (REML) was used for the longitudinal data because maximum likelihood (ML) presumes that the estimates are conditional on point estimates of the fixed effects. REML estimates adjust for uncertainty about the fixed effects.

First, Intraclass Correlation Coefficient (ICC) is checked to see the proportion of total variance in the outcome that lies between groups:

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} :$$

With the ICC equation, 46.53 % of the variability in writing gains is at the student-level. Larger ICC tends to increase the reliability because the between-group variance is large relative to within-group variability. As a result, it may be easier to discriminate between groups.

For the *fixed effects*, the average intercept, when TIME equals zero (i.e., the pre-test), for the students in the Control group (a reference group) is $\gamma_{00}=10.10$. For the students in the Intentional group (INTENTIONAL), the average intercept increased by 2.19 points ($\gamma_{01}=2.19$). For the students in the Incidental group (INCIDENTAL), the average intercept increased by 3.49 points ($\gamma_{01}=3.49$). The effect of the Incidental group was statistically different from zero ($t(47)=2.76$, $p=.008$); whereas, the one of the Intentional group was not different from zero ($t(47)=1.76$, $p=.086$). Overall, the average slope for the effect of TIME on lexico-grammatical gains, for the students in the Control group is $\gamma_{10}=1.46$. The effect of TIME on both groups were not statistically significant. In other words, the rate of growth was not different among the control and intervention groups.

Table 6. Final estimation of fixed effects

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. df.	p-value	Reliability
For INTRCPT1, β_0						0.511
INTRCPT2, γ_{00}	10.10	0.96	10.48	47	<0.001	
INTENTIONAL, γ_{01}	2.19	1.25	1.76	47	0.086	
INCIDENTAL, γ_{02}	3.49	1.26	2.76		0.008	
For TIME slope, β_1						0.019
INTRCPT2, γ_{10}	1.46	0.53	2.77	47	0.008	
INTENTIONAL, γ_{11}	0.54	0.68	0.79	47	0.435	
INCIDENTAL, γ_{12}	-0.43	0.69	-0.63	47	0.534	

For the *random effects*, the values in the random effects portion tell us about *variability* in the level-two estimates. The between-student residual variance in the intercepts (τ_{00}) is 6.17. It is important that the intercepts are the expected values when TIME is zero for the individual student. There is a lot of variability remaining in the intercepts since the chi-square test is statistically significant ($\chi^2=93.64$, $p<.001$). The between-student residual variance in the slopes from this model, $\text{Var}(u_{1j})=\tau_{11}=0.07$. Variability in the slopes across fifty students did not seem to be accounted for TIME because the chi-square test was not statistically significant.

Table 7. Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	df.	χ^2	p-value
INTRCPT1, u_0	2.48	6.17	47	93.64	<0.001
TIME slope, u_1	0.26	0.07	47	43.27	>0.500
level-1, r	2.66	7.09			

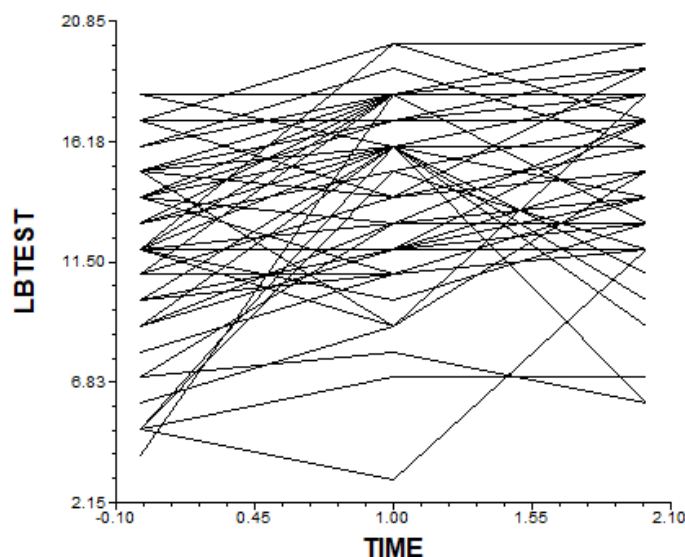


Figure 3. Change of the individual students' lexico-grammatical writing gains over time

Proportion Reduction in Error (PRE) values was checked to examine how much variability in the dependent variable of lexico-grammatical writing gains was accounted for by the model. This can be checked through the reduction in σ^2 and τ_{ij} values compared across competing models. 2.41 % of the variance within-students is accounted for by the inclusion of INTERVENTION. Slight variation ($\chi^2(4) = 13.06, p = .011$) is expected to occur because all the fixed and random effects are estimated jointly. 22.31 % of variance in student-intercepts was accounted for by INTERVENTION. There must be potential level 2 variables that can explain the lexico-grammatical writing gains because there is still variance in the student mean scores. Zero percent of variance (because of -5.3%) in student-slopes was accounted for by INTERVENTION. The p-value was over .5 in the conditional model for variance of the slopes, indicating that variability in the effect of TIME on writing gains was not explained by INTERVENTION. The proportional reduction in the prediction error variance for the time in level 1 (i.e., pseudo R^2) was checked with the equation below, and 13 % of variance in PRE level 1 was yielded. The proportional reduction in the prediction error variance for the intervention in level 2 was checked, and 22 % of variance in PRE level 2 was yielded.

The comparison between the results of repeated measures ANOVA and MLM

As expected, the results from Repeated Measures ANOVA and MLM were same based on the same dataset. However, the mechanism of approaching the dataset was different. Therefore, it is important for researchers to consider what kind of statistical methods should be employed in dealing with longitudinal data. The Repeated Measures ANOVA checks the assumptions of homoscedasticity and sphericity so that it can produce robust results. The Repeated Measures ANOVA requires fewer subjects if they meet the assumptions, such as linearity, independence, and homogeneity of variance. Using SPSS software was quick and user-friendly, providing efficiency and convenience. MLM is flexible in terms of meeting the assumptions (e.g., homoscedasticity and sphericity) although it did not matter in this dataset. MLM allows modeling of the variance-covariance matrix from the data. MLM can also be used to analyze multiple levels of data; whereas, Repeated Measures ANOVA is limited to repeated measures within subjects. The huge contribution

of using MLM is to obtain individual variability and the information about whether the rate of growth relies on personal characteristics in longitudinal data.

DISCUSSION

The major differences between Repeated Measures ANOVA and MLM is that MLM can also check the individual trend by posing two or multiple different levels. We can explore both individual level characteristic in Level 1 and contextual characteristics in Level 2. In this paper, time was added in Level 1, and the individual characteristic in Level 2 to analyze the longitudinal data. The second difference is that Repeated Measures ANOVA must have the same measure schedule; whereas, MLM does not require us to set up the same measurement schedule. MLM tolerates differently spaced waves of data from different subjects. Third, Repeated Measures ANOVA requires sphericity—the condition where the variances of the differences between all levels of the independent variable are equal, while MLM does not require it with the capability of analyzing incomplete data [21]. Overall, MLM is robust against homoschedasticity, sphericity, and missing data [22].

Repeated Measures ANOVA is also important in light of several advantages: statistical power, fewer subjects, and efficiency. As one of the parametric analyses, Repeated Measures ANOVA can be powerful because it can control for factors that cause variability between subjects. It can use fewer subjects to detect a desired effect size. Since fewer subjects may be sufficient for the analysis, it is quicker and cheaper to recruit, train, and compensate to complete an entire experiment. Furthermore, the format of the data (a wide format for the person data) in Repeated Measures ANOVA is relatively easier than the one (a long format for the person period) in MLM.

This investigation allows second language researchers to take into account their empirical data in an insightful way. This, in turn, enhances scientific inferences from a limited set of observations. Both Repeated Measures ANOVA and MLM have their own advantages and disadvantages. It seems that MLM requires less stringent assumptions than Repeated Measures ANOVA. Therefore, it is appealing to employ MLM without methodological considerations of its computational complexity and density. This advanced statistical technique is often overwhelming to second language researchers if they do not have any chance of methodological training or sufficient data to undertake it. It is essential to carefully and rigorously examine the characteristics of data (e.g., longitudinal data) and seek the adequate research analysis, particularly to find the significant effect of second language interventions that contributes to benefitting second language research.

References

1. C. E. McCulloch, G. Casella, and S. R. Searle. Variance components. New York, NY: Wiley, 1992.
2. J. Coady. L2 vocabulary acquisition: A synthesis of research. In J. Coady and T. Huckin (eds.) Second language vocabulary acquisition. Amsterdam: Benjamins, pp. 273–290, 1997.
3. S. Gass. Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), pp.319–333, 1999.
4. R. Schmidt. Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (ed.) Implicit and explicit learning of languages. London. England: Academic Press, pp. 165–210, 1994.
5. J. Hulstijn. Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (ed.) *Cognition and second language instruction*. New York, NY: Cambridge University Press, pp. 258–286, 2001.

6. J. Hulstijn. Incidental and intentional learning. In C. Doughty and M. H. Long (eds.) *The handbook of second language acquisition*. Malden, MA: Blackwell Publishing, pp. 351–381, 2003.
7. S. P. Nation. *Learning vocabulary in another language*. Cambridge: Cambridge University Press, 2001.
8. S. P. Nation and S. Webb. *Researching and analyzing vocabulary*. Boston, MA: Heinle Cengage Learning, 2011.
9. D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan. *Longman grammar of spoken and written English*. London: Pearson, 1999.
10. V. Cortes. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, pp. 397–423, 2004.
11. K. Hyland. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, pp. 4–21, 2008.
12. M. Kazemi S. Katiraei, and A. E. Rasekh. The impact of teaching lexical bundles on improving Iranian EFL students' writing skill. *Procedia-Social and Behavioral Sciences*, 98, pp. 864–869, 2014.
13. L. A. Carroll. *Rehearsing new roles: How college students develop as writers*. Carbondale and Edwardsville: Southern Illinois University Press, 2002.
14. J. Harris. The idea of community in the study of writing. *College Composition and Communication*, 40(1), pp. 11–22, 1989.
15. K. Hyland. *Academic discourse*. London: Continuum, 2009.
16. Leki. *Undergraduates in a second language: Challenges and complexities of academic literacy development*. Mahwah, NJ: Lawrence Erlbaum, 2007.
17. R. Gueorguieva, and J. H. Rogosa. Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61(3), pp. 310–317, 2004.
18. D. Rogosa, D. Brandt, and M. Zimowski. A growth curve approach to the measurement of change. *Psychological Bulletin*, 92, pp. 726–748, 1982.
19. S. W. Raudenbush. Toward a coherent framework for comparing trajectories of individual change. In A. G. Sayer (eds.) *New methods for the analysis of change*. Washington DC: American Psychological Association, pp. 35–64, 2001.
20. C. Teddlie and A. Tashakkori. *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, CA: SAGE, 2009.
- A. S. Bryk, and S. W. Raudenbush. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: SAGE, 1992.
21. S. W. Raudenbush and A. S. Bryk. *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE, 2002.
22. T, Teo. *Handbook of quantitative methods for educational research*. Springer Science & Business Media, 2014.