Effect Size for a Multilevel Model Random Slope Effect: Change in Variance Accounted for with Likelihood-based versus Variance Partition Measures

Julie Lorah
Indiana University, Bloomington

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Effect Size for a Multilevel Model Random Slope Effect: Change in Variance Accounted for with Likelihood-based versus Variance Partition Measures

Julie Lorah, *Indiana University, Bloomington*

Applied educational researchers may be interested in exploring random slope effects in multilevel models, such as when examining individual growth trajectories with longitudinal data. Random slopes are effects for which the slope of an individual-level coefficient varies depending on group membership, however these effects can be difficult to interpret. The change in variance accounted for is often used as an effect size measure and could be appropriate for helping to interpret a random slope effect. Two methods for computing variance accounted for include likelihood-based methods and variance partition methods. It is unclear how results from these two methods compare to each other when used to compute a measure of change in variance accounted for with a random slope effect. The present study fills this gap through a simulation study comparing these two methods under various conditions. Results indicate that the value of variance accounted for may differ depending on the type of measure used, and that applied researchers should consider reporting values for both measures.

Introduction

The multilevel model can be used to account for nesting in data, such as when students are nested within schools. Further, a random slope effect can be included in these models to allow the slope of an individual-level predictor to vary by group membership. This model is commonly used in education, for example, to measure academic growth with longitudinal data where observations at multiple time points are nested within students (Wright, 2017). Effect size measures for the fixed and random effects associated with a random intercept model are available for applied researchers.

However, currently there is very little guidance regarding what measure of effect size can be used for a random slope effect. This may be problematic for researchers since the random slopes model is commonly used in practice (Rights & Sterba, 2019)

while at the same time professional organizations and journals are more often requiring effect size reporting in addition to or in place of null hypothesis significance testing (Kelley & Preacher, 2012; Peng & Chen, 2013). In addition, effect size measures are considered an important reflection of the practical significance of findings and therefore represent a key finding for applied researchers; in addition to the fact that generally effect size measures are necessary for future researchers conducting power analyses or meta-analysis (Denson & Seltzer, 2011; Dong et al., 2020; Kelley & Preacher, 2012). Given these various potential uses, it is important for researchers to consider to what specific uses the effect size measure will contribute when choosing an appropriate effect size measure.

The random coefficient model (including a random slope) with a single predictor can be expressed as (Snijders & Bosker, 2012):

$$Y_{ij} = \beta_0 + \beta_1 * X_{ij} + U_{0j} + U_{1j} * X_{ij} + e_{ij}$$

$$Var(e_{ij}) = \sigma^2$$

$$Var(U_{0j}) = \tau_0^2 \qquad\qquad (1)$$

$$Var(U_{1j}) = \tau_1^2$$

$$Cov(U_{0j}, U_{1j}) = \tau_{01}$$

Where Y_{ij} represents the continuous outcome, β_0 represents the intercept, β_1 represents the slope for predictor X_{ij}, X_{ij} represents an individual-level predictor, U_{0j} and U_{1j} represent level-two residuals, e_{ij} represents level-one residuals, i represents individuals (level-one), and j represents group (level-two). Note that level-two residuals represent the difference between overall average and group-specific values and are not related to a difference between observed and predicted values in this context. This model can help answer important questions about whether and how the relationship between the independent and dependent variable vary by group membership. Note that random slopes are analogous to a fixed effect interaction effect, with the distinction that the interaction is now between a fixed and a random effect (Lorah, 2018).

To provide evidence of significance for random slopes, an analyst can conduct a mixture likelihood ratio test (Stram & Lee, 1994). However, once evidence has been provided for retaining the random slope effect, it may be difficult to interpret (Goldstein, Browne, & Rasbash, 2002). The random slopes are assumed to be normally distributed with mean β_1 and variance τ_1^2 which indicates that interpretation can proceed by computing specific values within this distribution, such as ±2 standard deviations from the mean. This computation would result in two slope values within which about 95% of the slopes exist (Snijders & Bosker, 2012). Further, the covariance term can be standardized and reported as a correlation coefficient which can be interpreted according to standard criteria and the random slopes effect can be plotted, as is typical for interaction effects (Lorah, 2018).

For a random intercept model, the intraclass correlation coefficient (ICC; ratio of between-cluster variance to total variance) can be used as an effect size measure (Lorah, 2018; Snijders & Bosker, 2012) and this value represents the correlation between individual observations and cluster membership (Leckie et al.,

2020). Alternatively, the variance partition coefficient (VPC) can be used to represent the proportion of outcome variance attributable to cluster membership. For the random intercepts model, the ICC and VPC are equal (Goldstein et al., 2002; Leckie at al., 2020).

However, for a model with random slopes where there is no longer a single source of variation at each level, the ICC and VPC are no longer equal, and the VPC becomes less useful for interpretation (Goldstein et al., 2002). However, since the VPC is a function of the predictor variable in the model, the VPC may still be computed for specific values of the associated predictor, or plotted across several values of the predictor to aid interpretation of the model. Goldstein et al. (2002) provide detailed instructions for doing so for the random slopes model, as well as several additional generalizations including discrete response multilevel models. More recently, researchers have extended these variance partition methods to various multilevel models, such as logistic models with overdispersion (Brown et al., 2005); and models examining count data (Leckie et al., 2020).

Although the VPC can be helpful for interpreting the proportion of variance attributable to cluster membership, methods for interpreting the random slope effect in particular are still needed. To do so, the researcher may choose to directly interpret the standard deviation of the random slopes (i.e. square root of τ_1^2; Lorah, 2018), or to compute a change in variance accounted for measure based on an appropriate variance accounted for value (Rights & Sterba, 2020). Variance accounted for could be computed based on a measure using the likelihood value or a measure based on variance partitioning; however, it is not clear how these two types of measures compare to each other and which one to report.

Literature Review

Effect size measures for random slopes models are made more complicated by the fact that partitioning variance is less straight-forward compared with the random intercepts model. This is clarified by examining the variance of Y_{ij}, which depends on the predictor, X_{ij} (Snijders & Bosker, 2012):

$$Var(Y_{ij} \mid X_{ij}) = \tau_0^2 + 2\,\tau_{01}\,X_{ij} + \tau_1^2 X_{ij}^2 + \sigma^2 \qquad (2)$$

Note also that the value of $\tau_0{}^2$ will vary depending on how the predictor X is centered (Snijders & Bosker, 2012)

Although the literature provides various variance accounted for measures specifically for random slopes models (Johnson, 2014; Rights & Sterba, 2019), some methodologists argue that when evaluating a random slopes model, the variance accounted for value for the analogous random intercepts model with the same fixed effect should be used (Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1994). This recommendation is supported by the claim that computation of variance accounted for within the random slopes model may be tedious (Nakagawa & Schielzeth, 2013); that the values should be quite similar (Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1994); that since the level-2 residuals (U_{0j} terms) are unknown, they do not help predict the outcome Y_{ij} (Snijders & Bosker, 1994); and because the contribution of the random slopes to predicting outcome variance is typically small (LaHuis et al., 2014).

Based on a simulation study, LaHuis et al. (2014) find that the values for variance accounted for in the random intercepts model are similar to those for the random slopes model when the slope variance is small, but do not provide a good approximation when the slope variance is large. Johnson (2014) suggests that the correspondence between the two measures depends on how accurately the random intercepts model estimates the slope coefficient as well as whether the number of observations within groups is relatively similar or not. Therefore, depending on the nature of the data, the difference in variance accounted for between the random intercept and random slopes model has the potential to be considerable (Johnson, 2014).

In order to measure the unique contribution of an effect, the change in R^2 value can be computed based on models with and without the given effect (Darlington & Hayes, 2017) which is used in the present study. Note that the present study considers R^2 measures derived from variance, as in OLS models, as well as R^2 measures derived from the value of deviance, which is distinct from those used for OLS models.

Likelihood-based Measures

The first measure considered is (McFadden, 1974):

$$R_{MF}^2 = 1 - \frac{\ln(L_M)}{\ln(L_0)} \qquad (3)$$

where ln indicates the natural log; L_M is the value of the model's likelihood function and L_0 is the value of the baseline model's likelihood function.

This measure can be considered to be based on "deviance decomposition" analogous to the variance decomposition achieved by the OLS measure (Veall & Zimmerman, 1996). This measure also represents proportional reduction in error (Menard, 2000) and has been examined in the context of logistic regression (Menard, 2000) and limited dependent models (Veall & Zimmerman, 1996) although no instance of examining this measure for use with multilevel models was found. Menard (2000) recommends this measure due to its intuitive interpretation. Further, Menard (2000) compared this measure with the OLS measure (based on variance) and based on seven empirical analyses found the average value for the McFadden measure was about .23 whereas for the OLS measure it was about .19 based on a logistic regression model.

Variance Partition Measures

The next measure considered is based on a partition of variance appropriate for the random slopes model (Johnson, 2014). The total variance is comprised of variance due to fixed effect, random effects, and residual variance. Based on the variance partition (VP), the conditional measure of variance accounted for can be defined as:

$$R_{VP}^2 = \frac{\sigma_f^2 + \Sigma\sigma_l^2}{\sigma_f^2 + \Sigma\sigma_l^2 + \sigma_e^2} \qquad (4)$$

where σ_f^2 is the variance of the fixed effects, σ_l^2 is the variance of the lth random effect, and σ_e^2 is the residual variance. Within the random slopes model, the value for $\Sigma\sigma_l^2$ is dependent on the value of the predictor, X_{ij}. Therefore, in order to compute this value when random slopes are included, the following provides the mean random effect variance across all observations X_{ij} (Johnson, 2014):

$$\Sigma\sigma_l^2 = Tr(Z\Sigma Z')/n \qquad (5)$$

where Z represents the design matrix for the random effects, Σ represents the covariance matrix of random effects, Tr represents the matrix trace operation, Z' represents the transpose of Z, and n represents the number of observations. The design matrix, Z, would be represented by a column of the value one (for the random intercept) and a second column of the values of X_{ij} for the basic random slopes model with a single predictor described by equation 1. Note that equation 5 is needed with the random slopes model due to the dependence of the random effect variance component on X_{ij} (Johnson, 2014).

This method for computing variance accounted for has been described for the generalized linear multilevel model with random intercepts only (Nakagawa & Schielzeth, 2013), and within a random intercepts and random slopes multilevel model for both the linear and the generalized linear models (Johnson, 2014) and evaluated empirically through simulation study based on the linear random intercepts and random slopes multilevel model (LaHuis et al., 2014). Further, Rights and Sterba (2019) generalize this framework to provide measures that can, for example, differentiate among variance accounted for from level-one fixed effects versus level-two fixed effects and explore R^2 difference measures appropriate for various effects (Rights & Sterba, 2020). Note that Rights and Sterba (2019) show that their measure is a generalization of the Johnson (2014) measure (see Rights and Sterba (2019), Appendix Section B3). Although this generalization allows the researcher a bit more flexibility, the Johnson (2014) measure was chosen for the present study as it does not require group-mean centering of variables for any variation of the measure, the associated software option is a bit more user-friendly, and it is specifically derived and offered just for random slopes effect, the focus of the present study.

This measure can be computed automatically from the r.squaredGLMM function from the MuMIn package (Barton, 2019) which is available in R (R Core Team, 2021), and was the method used in the present study.

Marginal versus Conditional Measures

Measures of variance accounted for in a multilevel model are often classified as either conditional or marginal measures. Note that this distinction refers to the measures themselves, not the difference measures. Conditional measures represent variance accounted for by both the fixed and random effects of the model, while marginal measures represent variance accounted for by fixed effects only (Nakagawa & Schielzeth, 2013; Orlien & Edwards, 2008). Given that the present study examines random effects specifically, conditional measures must be used so that variance accounted for by the random slope effect is included in the measure. This implies that the baseline model for the McFadden measure should be the OLS intercept-only model, in order for the measure to include variance explained due to both the fixed and random effects.

Evaluation Criteria

The change in R^2 values between a model with and without a random slope effect are evaluated based on the following criteria:

1. The values should not be related to sample size. Specifically, Kelley and Preacher (2012, pg. 147, property number 3) indicate that a good effect size measure should be independent of sample size.

2. The values should be highly related to the true value of the effect, τ_1^2. In order for the measure to have "utility as a measure of goodness of fit and an intuitively reasonable interpretation" (Kvalseth, 1985, pg. 281, criteria 1), the measure should clearly be related to the effect which it is describing.

3. The values should not be negative. A negative change value would indicate that the variance accounted for has decreased when a random slopes effect was added. However, criteria for measures of variance accounted for indicate that these values should not decrease whenever effects are added (Cameron & Windmeijer, 1996, criteria 2).

4. The values upon repeated replication should not vary widely. Low variability among replications indicates an efficient estimator and this can be measured through the standard deviation of repeated replications, which approximates the standard error (LaHuis et al., 2014).

Present Study

The present study compares two potential measures of effect size for a random slopes effect. Both measures are computed as the difference in variance accounted for between a model with a single random slope effect (equation 1) and an identical model with the random slope effect removed. The measures of variance accounted for used are:

1. McFadden. Computed as specified in equation 3 with the baseline represented by the OLS intercept-only model.

2. Variance partition. Computed as specified in equation 4 with random effect variance as specified in equation 5.

The following research question is investigated: When estimating a random slopes model, how does a likelihood-based measures of variance accounted for compare with a variance partition measure for computing a measure of change in variance accounted for as an effect size measure of a random slope effect?

Methods

To evaluate these possibilities for effect size, Monte Carlo simulation was used in R (R Core Team, 2021) and code for the simulations is provided in the appendix. The parameter values for the simulation study were chosen based on the literature and specific goals of this study. Specifically, the present study uses 2000 simulations per conditions, slightly more than previous simulation work examining variance accounted for with multilevel models which used 500 (Rights & Sterba, 2019). Previous multilevel modeling simulation work examining variance accounted for has used a sample size of 200 groups with 50 observations per group (Rights & Sterba, 2019) and 40, 70, and 160 groups with average group size of 4, 8, and 21 (LaHuis et al., 2014). Additionally, this simulation work has used a value of 17 for residual variance (σ^2) and values of 1, 1.5, 2, and 10 for the variance for random effects (Rights & Sterba, 2019) as well as τ_{00}=.176 or .429 and τ_{01} value of 0.1 (LaHuis et al., 2014) and ICC of .15 and .30 (LaHuis et al., 2014).

Based on these conditions just summarized from the literature, four parameters were varied and fully crossed in the present study including group size (values of 10, 40, and 50); number of groups (values of 20, 50, and 70); τ_0^2 (values of 6.25, 9, and 12.25); and τ_1^2 (values of 1, 2.25, 4, and 6.25) resulting in a total of 108 conditions. Number of groups represents the sample size at level two while the group size represents the sample size within each level-two group. The conditions in the present study resulted in an observed ICC that varied by condition and ranges from about 0.18 to about 0.33 based on the simulated datasets. Note that additional analyses were run varying τ_{01} (with values of 0, 0.3, and 0.6), but no relationship between this covariance and either of the measures of variance accounted for was found (all correlations around 0.02 or lower) and so for the sake of clarity, τ_{01} was held constant for the final analyses. For each condition, 2000 simulations were run. Data was generated according to equation 1 and β_0 was held constant at five; β_1 was held constant at two; and τ_{01} was held constant at zero. The individual-level predictor X_{ij} was generated as a standardized random normal variable with mean of zero and standard deviation of one. The individual-level residual, e_{ij} was generated as a random normal variable with mean of zero and variance of 16.

The following two models were estimated (all notation consistent with equation 1):

Empty OLS model: $Y_{ij} = \beta_0 + e_{ij}$ (6)

Random intercept model:

$Y_{ij} = \beta_0 + \beta_1 * X_{ij} + U_{0j} + e_{ij}$ (7)

along with the random slopes model (equation 1) for a total of three models. All data was generated and analyzed using R (R Core Team, 2021).

To analyze the simulated data, both of the measures described above were computed for the random intercepts model (equation 7) and for the random slopes model (equation 1) for each replication. The difference between these two values was used as an effect size measure for the random slope effect. To assess these measures, correlations between the two measures, as well as among the measures and parameters were computed. In addition, the minimum value, maximum value, mean, and standard deviation for the values for each measure were computed. Note that although it would be informative to compare these measures to their true, population value through metrics such as bias and RMSE, this is not possible in the present study because the model parameters are

Lorah, Random Slopes Effect Size

Table 1. Results for Both Change Accounted for Measures

	McFadden	Variance Partition
Correlations		
1. Group size	0.293	0.010
2. # of groups	0.013	0.025
3. Total N	0.223	0.021
4. Intercept variance	-0.016	-0.110
5. Slope variance	0.808	0.844
Min	0.000	-0.012
Max	0.104	0.526
Mean	0.018	0.098
SD	0.014	0.062

Note. Both measures computed based on change in variance accounted for between a model with and without the random slope effect. Results based on 2000 simulations for each of 108 conditions. The first five rows represent the correlations between change in variance accounted for values and the simulated parameter conditions.

specified within the simulation but not the variance accounted for measures.

Results

Table 1 provides correlations among measures of change in variance accounted for and parameters as well as the minimum value, maximum value, mean, and standard deviation for each measure.

Sample Size

Rows 1-3 of Table 1 show the correlations between each of the measures and three aspects of sample size (Group size, # of groups, and Total N). The McFadden measure is moderately positively correlated with group size and overall sample size, while the variance partition measure shows negligible correlation with sample size. Generally speaking, one advantage of effect size reporting compared with hypothesis testing is that effect size measures are not influenced by sample size (Kelley & Preacher, 2012). In this case, the fact that the McFadden measure shows small positive correlations with group size may be considered problematic. Based on this metric, and consistent with evaluation criteria 1 specified above, the variance partition measure would be preferred.

Intercept Variance

The correlations between both measures and the intercept variance value can be found in Table 1, row 4. Both measures show small negative correlation with the intercept variance, with the variance partition measure indicating a slightly stronger correlation. This is expected. As intercept variance (τ_0^2) increases, this causes the total variance to increase. With slope variance (τ_1^2) held constant and total variance increasing, the proportion of variance attributable to

slope variance decreases. This is demonstrated with the negative correlation between intercept variance and variance accounted for by the random slope.

Slope Variance

The correlations between both measures and the slope variance value can be found in Table 1, row 5. Slope variance is strongly positively correlated with both measures, which is expected since this is the parameter manipulating the effect size of the random slopes. Thus, according to evaluation criteria 2, which specifies that the value should be highly related to the true effect, both measures may be used.

Comparison Among Measures

These two measures are correlated at 0.909, indicating a high degree of overlap with each other.

According to Table 1, the minimum value for the McFadden measure is zero indicating that the inclusion of the random slopes effect didn't cause a decrease in the variance accounted for value for any simulation replications in this study. For the variance partition measure, the minimum value is -0.012, indicating that in some simulations, including the random slopes effect caused a small decrease in the variance accounted for value. Ideally, the change in variance accounted for when adding an effect won't be negative (Cameron & Windmeijer, 1996), indicating that based on this outcome and according to evaluation criteria 3 specified above, the McFadden measure is preferred.

The average value for change in variance accounted for differs for the different measures. The McFadden measure shows a much lower average value of 0.018 compared with the average value for the variance partition measure of 0.098.

The results and interpretation based on these different measures varies widely. For example, depending on the measure selected, this could indicate anywhere from about 2% to about 10% of variance in the outcome is accounted for by the random slope effect. This result calls into question the reasonableness of reporting a single measure of change in variance accounted for with a random slope effect based on evaluation criteria 2 which indicates the measure should have an intuitively reasonable interpretation (Kvalseth, 1985).

The standard deviation of these change measures across replications can be considered as a standard error and represent a measure of efficiency for the estimators (LaHuis et al., 2014). The standard deviation for the McFadden measure is 0.014, while for the variance partition measure it is 0.062. A smaller standard deviation value represents a more efficient estimator, which is preferable. In the present study, the McFadden measure displayed the smallest standard deviation value and may be preferred due to the higher efficiency.

Discussion & Conclusions

Both types of measures examined demonstrated desirable properties. The McFadden measure was not highly correlated to intercept variance, highly correlated to slope variance, did not take any negative values, and demonstrated high efficiency. On the other hand, the variance partition measure was not highly correlated to sample size, demonstrated slightly higher correlation with intercept variance, and was highly correlated with slope variance. Although the correlation between the two measures was 0.909, they demonstrated somewhat different average values at around 0.018 for McFadden and 0.098 for the variance partition measure. It's not clear why these average differences emerged, but this could be related to the fact that the random effect variance component depends on the predictor, X (Johnson, 2014) or possibly due to the fact that the level-2 residual (u_{0j} terms) are unknown (Snijders & Bosker, 1994). Menard (2000) did not find that the McFadden measure produced consistently lower values than a measure based on variance for a different model, indicating that the direction of this difference may be depend on the specific models and data examined. Consistent with results related to the logistic regression model (Menard, 2000), this study finds that there is no clear reason to choose one of these two types of measures over the other. Therefore, in order to give a fuller picture of the results, it is recommended that both measures be reported for applied studies.

The present study is not without limitations. The conclusions based on any simulation study are limited to the specific conditions examined. Although efforts were made to ensure realistic conditions, future research may want to expand beyond these conditions and specifically include models with additional fixed effects including interaction and cross-level interaction

effects, additional random effects, and non-zero intercept slope covariance values. Future research should additionally propose alternative measures for effect size for random slope effects based on statistics other than change in variance accounted for. Further, future research should consider level-specific measures for assessing variance accounted for by random slopes.

This study fills a gap in the literature by offering a comparison through Monte Carlo simulation of likelihood-based versus variance partition methods of computing a value for variance accounted for, in order to compute an effect size measure for a random slopes effect. Results indicate that both measures perform well and that it is worthwhile for applied researchers to compute and report both types of measures in empirical studies.

References

Barton, K. (2019). MuMIn: Multi-Model Inference. R package version 1.43.15. https://CRAN.R-project.org/package=MuMIn.

Brown, W. J., Subramanian, S. V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society, Series A*, 168, 599-613.

Cameron, A. C. & Windmeijer, F. A. G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics, 14*(2), 209-220.

Darlington, R. B., & Hayes, A. F. (2017). Regression analysis and linear models: Concepts, applications, and implementation. New York, NY: Guilford.

Denson, N., & Seltzer, M. H. (2011). Meta-analysis in higher education: An illustrative example using hierarchical linear modeling. *Research in Higher Education, 52*, 215–244.

Dong, N., Kelcey, B., & Spybrook, J. (2020). Design considerations in multisite randomized trials to probe moderated treatment effects. *Journal of Educational and Behavioral Statistics.* Advance online publication. https://doi.org/10.3102/1076998620961492 .

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics, 1*(4), 223-231. https://doi.org/10.1207/S15328031US0104_02 .

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's RGLMM2 to random slopes models. *Methods in Ecology and Evolution, 5*, 944–946. http://dx.doi.org/10.1111/2041-210X.12225

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*, 137–152. https://doi.org/10.1037/a0028086.

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods, 17*(4), 433-451.

Leckie, G., Browne, W. J., Goldstein, H., Merlo, J., & Austin, P. C. (2020). Partitioning variation in multilevel models for count data. *Psychological Methods, 25*(6), 787–801. http://dx.doi.org/10.1037/met0000265.

Lorah, J. A. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-scale Assessments in Education, 6*(8). DOI: https://doi.org/10.1186/s40536-018-0061-2 .

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). Academic Press.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician, 54*, 17-24.

Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133-142. https://10.1111/j.2041-210x.2012.00261.x.

Orlien, J. G. & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R2 statistics. *Computational Statistics & Data Analysis, 52*, 1896-1907.

Peng, C. Y. J., & Chen, L. T. (2013). Beyond Cohen's d: Alternative effect size measures for between-subject designs. *Journal of Experimental*

Education, 82, 22–50. https://doi.org/10.1080/00220 973.2012.745471.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rights, J. D. & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309-338.

Rights, J. D. & Sterba, S. K. (2020). New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivariate Behavioral Research, 55*(4), 568-599. https://doi.org/10.1080/00273171.2019.1660605.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling. Thousand Oaks: Sage Publishing.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50*(4), 1171-1177.

Veall, M. R. & Zimmerman, K. F. (1996). Psudo-R^2 measures for some common limited dependent variable models. *Journal of Economic Surveys, 10*, 241-260.

Wright, D. B. (2017). Some limits using random slope models to measure academic growth. *Frontiers in Education: Assessment, Testing, and Applied Measurement, 9*, https://10.3389/feduc.2017.00058.

Citation:

Lorah, J. (2022). Effect Size for a Multilevel Model Random Slope Effect: Change in Variance Accounted for with Likelihood-based versus Variance Partition Measures. *Practical Assessment, Research & Evaluation*, 27(9). Available online: https://scholarworks.umass.edu/pare/vol27/iss1/9/

Corresponding Author:

Julie Lorah
Indiana University
Bloomington, IN, USA

Email: jlorah [at] iu.edu

Appendix A. R Code


```
####################
####GENERATE DATA####

library(lme4)          #for the lmer function
library(MuMIn)         #for the r.squaredGLMM function
nsim<-2000             #number of simulations per condition
numColumns_result<-27     #number of results quantities to store

#values for varying conditions
values_NL1<-c(10,40,50)  #values for L1 N (group size)
values_L2SD<-c(2.5,3,3.5)      #values for tau (SD of U0j)
values_NL2<-c(20,50,70)  #values for L2 N (# of groups)
values_SlopeSD<-c(1,1.5,2,2.5)      #values for the SD of U1j

#values for constants
b0<-5
b1<-2

#number of conditions for each parameter varied
ncond_NL1<-length(values_NL1)
ncond_L2SD<-length(values_L2SD)
ncond_NL2<-length(values_NL2)
ncond_SlopeSD<-length(values_SlopeSD)

#total number of conditions
ncond<-ncond_NL1*ncond_L2SD*ncond_NL2*ncond_SlopeSD
#total number of rows for result matrix
numRows<-nsim*ncond

#create matrix of conditions
numColumns_conditions<-4 #number of conditions that vary
conditions<-matrix(ncol=numColumns_conditions,nrow=numRows)
colnames(conditions)<-c("NL1","L2SD","NL2","SlopeSD") #condition names
conditions[,1]<-rep(values_NL1,length.out=numRows)
conditions[,2]<-rep(values_L2SD,each=ncond_NL1,length.out=numRows)
conditions[,3]<-
    rep(values_NL2,each=ncond_NL1*ncond_L2SD,length.out=numRows)
conditions[,4]<-
    rep(values_SlopeSD,each=ncond_NL1*ncond_L2SD*ncond_NL2,length.out=num
    Rows)

#create matrix to store simulation results
result<-matrix(ncol=numColumns_result,nrow=numRows)
```

```
#run nsim simulations
for(i in 1:(numRows)){
NL1<-conditions[i,1]
L2SD<-conditions[i,2]
NL2<-conditions[i,3]
SlopeSD<-conditions[i,4]
N<-NL1*NL2

#generate variables
X1<-rnorm(N)
ID<-rep(seq(from=1,to=NL2,by=1),times=NL1)
U0j<-rep(rnorm(NL2,mean=0,sd=L2SD),times=NL1)
U1j<-rep(rnorm(NL2,mean=0,sd=SlopeSD),times=NL1)
eij<-rnorm(N,sd=4)

#generate Y
Y<-b0+b1*X1+U0j+U1j*X1+eij

#estimate models
M0<-lm(Y~1)
M1<-lmer(Y~(1|ID),REML=F,control=lmerControl(optCtrl=list(xtol_abs=1e-8,
     ftol_abs=1e-8)))
M2<-lmer(Y~X1+(1|ID),REML=F,control=lmerControl(optCtrl=list(xtol_abs=1e-
     8, ftol_abs=1e-8)))
M3<-lmer(Y~X1+(X1|ID),REML=F,control=lmerControl(optCtrl=list(xtol_abs=1e-
     8, ftol_abs=1e-8)))

#save relevant quantities in "temp"
temp<-c(
mean(Y),   #Ybar
logLik(M0),    #loglik_M0
logLik(M1),    #loglik_M1
logLik(M2),    #loglik_M2
logLik(M3),    #loglik_M3
sum((Y-Ybar)^2),    #SSR
sum((Y-Ypred1)^2),  #SSM1
sum((Y-Ypred2)^2),  #SSM2
sum((Y-Ypred3)^2),  #SSM3
var(predict(M1,type="response")),  #varYhat1
var(predict(M2,type="response")),
var(predict(M3,type="response")),
as.data.frame(VarCorr(M1))$vcov[1],    #tau-squared, M1
as.data.frame(VarCorr(M1))$vcov[2],    #sigma-squared, M1
as.data.frame(VarCorr(M2))$vcov[1],    #tau-squared, M2
as.data.frame(VarCorr(M2))$vcov[2],    #sigma-squared, M2
```

```
as.data.frame(VarCorr(M3))$vcov[1],      #tau-squared, M3
as.data.frame(VarCorr(M3))$vcov[2],      #slope var, M3
as.data.frame(VarCorr(M3))$sdcor[3],     #int/slope corr, M3
as.data.frame(VarCorr(M3))$vcov[4],      #sigma-squared, M3
cor(Y,Ypred1)^2,                   #correlation between observed & pred. Y,
      squared
cor(Y,Ypred2)^2,
cor(Y,Ypred3)^2,
r.squaredGLMM(M2)[1],    #R2 marginal
r.squaredGLMM(M2)[2],    #R2 conditional
r.squaredGLMM(M3)[1],
r.squaredGLMM(M3)[2]
)
#add values in temp to full table of results
result[i,]<-temp
}

#assign names to the columns of "result"
colnames(result)<-
      c("Ybar","loglik_M0","loglik_M1","loglik_M2","loglik_M3",
      "SSR","SSM1","SSM2","SSM3",
      "varYhat1","varYhat2","varYhat3",
      "tau2M1","sigma2M1","tau2M2","sigma2M2",
      "tau2M3","slopeM3","corrM3","sigma2M3",
      "Cor1","Cor2","Cor3",
      "R2_M2_Mar","R2_M2_Cond","R2_M3_Mar","R2_M3_Cond")

#combine conditions and result
mydata<-as.data.frame(cbind(conditions,result))



#######################
####COMPUTE MEASURES####

# McFadden
mydata$McFadden<-(1-(mydata$loglik_M3/mydata$loglik_M0)) - (1-
      (mydata$loglik_M2/mydata$loglik_M0))

# Variance partition
mydata$Variance.Partition<-mydata$R2_M3_Cond-mydata$R2_M2_Cond
```