Comparing Between- and Within-Group Variances in a Two-Level Study: A Latent Variable Modeling Approach to Evaluating Their Relationship

Educational and Psychological
Measurement

2017, Vol. 77(2) 351–361

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164416634166

journals.sagepub.com/home/epm

Tenko Raykov¹, George A. Marcoulides², and
Hope O. Akaeze¹

Abstract

This note is concerned with examining the relationship between within-group and between-group variances in two-level nested designs. A latent variable modeling approach is outlined that permits point and interval estimation of their ratio and allows their comparison in a multilevel study. The procedure can also be used to test various hypotheses about the discrepancy between these two variances and assist with their relationship interpretability in empirical investigations. The method can also be utilized as an addendum to point and interval estimation of the popular intra-class correlation coefficient in hierarchical designs. The discussed approach is illustrated with a numerical example.

Keywords

between-group variance, confidence interval, hierarchical design, latent variable modeling, multilevel study, variance ratio, within-group variance

¹Michigan State University, East Lansing, MI, USA

²University of California, Santa Barbara, CA, USA

Corresponding Author:

Tenko Raykov, Measurement and Quantitative Methods, Michigan State University, 443A Erickson Hall,
East Lansing, MI 48824, USA.

Email: raykov@msu.edu

Multilevel studies have become very popular in the educational and psychological sciences over the past two decades (e.g., Goldstein, 2011). A main reason for this pronounced trend is the widespread realization that data collected in behavioral and social research are far more often of a hierarchical nature than previously assumed. In such investigations, however, the traditional assumption of independence of the units of analysis—an essential premise of conventional single-level statistical methods—cannot be considered fulfilled in general, because the characteristic feature of such studies is that subjects are clustered or nested within higher order units. As a result, these studied persons may not provide independent scores on response variables of interest. If the independence assumption is violated to a considerable degree then, standard methods based on it may yield seriously misleading results (e.g., Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002).

In two-level models, which represent the majority of models currently used in the educational and behavioral disciplines with hierarchical data, the degree of this lack of independence of subjects' outcome scores can be evaluated with the popular intra-class correlation coefficient (ICC). The ICC can be readily point and interval estimated, for instance, using a latent variable modeling approach (e.g., Raykov, 2011; see also, e.g., Raykov & Marcoulides, 2015, for the discrete response case), and three-level counterparts of the ICC can be similarly evaluated with suitable modifications of that approach (e.g., Raykov, 2010; Raykov, Patelis, Marcoulides, & Lee, 2016). While the ICC provides rather important information about the relationship of between-cluster to total variance in a two-level study, its empirical value is in general not easily interpretable by a practitioner. In particular, being a proportion between 0 and 1, or alternatively expressed as a percentage of total variance due to between group differences, the ICC is not always easy to interpret in substantive terms, especially when specific questions about the relationship of between-cluster to within-cluster variance rather than total variance are also of concern or even of more interest and importance to answer.

The present note provides a widely applicable method for enhancing the practical interpretability of the well-known variance decomposition strategy in hierarchical data sets. A latent variable modeling (LVM; e.g., B. O. Muthén, 2002) procedure is first outlined that permits routine point and interval estimation of the ratio of between- to within-group variance and complements the ICC evaluation in two-level settings. With this method, an empirical scientist can assess in arguably more pragmatically oriented terms the extent to which between-cluster variance is larger, or conversely is smaller, than within-cluster variance. In this way, he or she can utilize an alternative to relating the between-cluster variance to its sum with the within-cluster variance in the ICC, which could therefore be seen as confounding the two variances in its denominator. When one is interested in the extent to which between variance outperforms within variance, or vice versa, a researcher can thus evaluate directly their relationship and interpret straightforwardly in terms of relative size these two essential variances in a two-level study. We illustrate subsequently the

proposed procedure with a numerical example, and conclude with a discussion of its limitations in empirical research.

Background, Notation, and Assumptions

In the remainder of this article, we denote by Y_{ij} the score on a response variable under consideration for the i th subject (level-1 unit) in the j th group (level-2 unit; $i = 1, \dots, n_j, j = 1, \dots, J$; cf., e.g., Raudenbush & Bryk, 2002). In educational and behavioral research, level-1 units are frequently students, patients, employees or repeated measurements, which are nested (clustered) within level-2 units such as schools, treatment centers, companies, departments, or persons longitudinally followed across several measurement occasions. We assume in the sequel that Y_{ij} is an (approximately) continuous outcome measure (see also Conclusion section).

At the outset of a multilevel modeling process, the (fully) unconditional two-level model is defined by the next two equations that represent its within-group and between-group parts (models), respectively:

$$Y_{ij} = \beta_{0j} + r_{ij}, \quad (1)$$

and

$$\beta_{0j} = \gamma_{00} + u_{0j}. \quad (2)$$

In this unconditional model, r_{ij} is the deviation of Y_{ij} from the mean β_{0j} of the j th group, with this difference assumed to be normally distributed with mean 0 and variance σ^2 ($i = 1, \dots, n_j, j = 1, \dots, J$). The variance σ^2 is typically referred to as within-group (within-cluster) variance and is assumed to be positive in this article. In the second equation of the model, u_{0j} is the j th group's mean deviation from the grand mean γ_{00} , which discrepancy is also assumed normal, with zero mean and variance τ_{00} ($j = 1, \dots, J$). The variance τ_{00} is usually referred to as between-group (between-cluster) variance.

By substitution, the associated combined or mixed model is

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (3)$$

($i = 1, \dots, n_j, j = 1, \dots, J$). Equation (3) presents the basic decomposition used to obtain the popular ICC, which is defined as the ratio of between-group variance to its sum with within-group variance:

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2). \quad (4)$$

As seen by its definition, the ICC represents the proportion (percentage after multiplying by 100) of observed variance that is due to between-cluster differences, and contains rather useful information in many empirical circumstances. Given that the sum of between and within group variance features in the denominator of the ICC,

one may argue that (the denominator of) the ICC “confounds” between- and within-cluster differences. This property of the ICC may thus be seen by a practitioner as somewhat distracting, who may actually be more interested in the extent to which between-group differences surpass or are larger than within-group differences, or vice versa. For example, an educational policy analyst may be concerned with evaluating the extent to which school differences may be overwhelming student differences within each school, or alternatively may be interested in evaluating the degree to which within-cluster differences surpass between-cluster variability, with important policy related implications in either situation. In that case, the ratio of these two variances, defined as

$$r = \tau_{00} / \sigma^2, \quad (5)$$

can be seen as being of additional or possibly even greater relevance to him or her than the ICC.

While the between-to-within variance ratio r can be obtained from the ICC via corresponding algebraic rearrangements, the availability of a procedure to automate the calculation and provide point *and* interval estimation of this ratio would be quite helpful to empirical researchers. In addition, there do not seem to be widely available rules-of-thumb that could be used to evaluate in substantive terms an arbitrary ICC estimate in a given study; with this in mind, one might argue that a sample ICC estimate is in general not readily interpretable. By way of contrast, an estimate of the ratio (Equation 5) directly represents the number of times between-group variance is larger, or conversely smaller, than within-group variance. With this feature, the quantity r may be viewed by some applied researchers of additional if not greater value in certain circumstances than the ICC. We describe next a widely applicable procedure that readily accomplishes point and interval estimation of the ratio r in Equation (5).

A Latent Variable Modeling Procedure for Evaluation of the Ratio of Between-Group to Within-Group Variance in Two-Level Models

Point and Interval Estimation

Revisiting the unconditional model in Equations (1) and (2), or its mixed model version in Equation (3), we notice that it involves two latent variables (cf. Raykov, 2011). Indeed, none of the variables u_{0j} or r_{ij} appearing in the right-hand side of Equation (3) is measured or observed, unlike the recorded scores Y_{ij} ($i = 1, \dots, n_j$, $j = 1, \dots, J$). In addition, each corresponding level-2 or level-1 unit is associated with an individual realization of the random variable u_{0j} or r_{ij} ($i = 1, \dots, n_j$, $j = 1, \dots, J$). Hence, one can view Equations (1) and (2) as defining a latent variable model. This model can be readily fitted to a given empirical data set using popular LVM software, such as *Mplus* (L. K. Muthén & Muthén, 2015).

This LVM application will yield estimates of the between and within group variances as well as associated standard errors and covariances, and hence a point

estimate of the ratio (Equation 5) of interest in this article. Subsequently using the delta method (e.g., Raykov & Marcoulides, 2004), a confidence interval (CI) for this ratio r , for any prespecified confidence level can be obtained. More specifically, because Equation (5) involves two variances, r can never be negative, that is, the variance ratio r is a population parameter that is bounded from below (i.e., 0 is its lower bound).¹

Therefore, to obtain a CI for r at a given confidence level, say $(1 - \alpha)100\%$ ($0 < \alpha < 1$), we can utilize a monotone transformation approach (e.g., Browne, 1982). Accordingly, one needs to take first the (natural) logarithm of the estimator of r , $\log(r)$, and then using its standard error (obtained with the delta method) furnish a symmetric $(1 - \alpha)100\%$ CI by subtracting and adding $z_{\alpha/2}$ times that standard error to $\log(r)$, with $z_{\alpha/2}$ being the $(1 - \alpha/2)$ th standard normal quantile. Finally, via exponentiation of the last CI's limits, we obtain the same confidence level CI limits of the critical ratio r of initial interest. (Further discussion and details on this approach to CI construction are provided in Raykov & Marcoulides, 2014, in an analysis of incomplete data context, and we refer to that source.) All these activities leading to a CI for the variance ratio r , are implemented in the R-function “ci.r” provided in the appendixes to this note along with the *Mplus* source code needed to fit the two-level model (1) and (2) (cf. Raykov, 2011).

We illustrate in a subsequent section this point and interval estimation procedure using an empirical example.

Testing Simple and Composite Hypotheses About the Ratio of Between to Within Variance

While the ratio (Equation 5) is nonnegative and in practical terms could be seen as always positive (see Note 1), it may be of relevance in some settings to test the null hypothesis of it being 0, that is,

$$H_0 : r = 0 \quad (6)$$

In those cases, we mention in passing that the test of this hypothesis is in fact implemented in popular software, for instance *Mplus*, since null hypothesis (6) is equivalent to the hypothesis of vanishing between-group variance, that is, $H_0^* : \tau_{00} = 0$. The two-tailed p -value for the latter hypothesis is routinely provided by the software used, and dividing it by 2 renders the sought p -value for the null hypothesis (6), owing to the fact that formally the alternative hypothesis, $H_1^* : \tau_{00} > 0$, is one-tailed (e.g., Rabe-Hesketh & Skrondal, 2012).²

Further, for any given number $r_0 > 0$, which is of substantive interest to a researcher, a test of the hypothesis

$$H_0 : r = r_0 \quad (7)$$

at a prefixed significance level α against the bidirectional alternative of r not being equal to r_0 , is directly carried out by checking if r_0 is contained in the corresponding

CI obtained as above, with no rejection of hypothesis (7) in that case and rejection otherwise (e.g., Raykov & Marcoulides, 2008).

If the hypothesis

$$H_0 : r \geq r_0 \quad (8)$$

is to be tested against the alternative

$$H_0 : r < r_0, \quad (9)$$

at the significance level α , one can proceed in a similar way. Specifically, if the corresponding CI is entirely above r_0 , then H_0 can be considered not rejected; if that CI is entirely below r_0 , H_0 can be rejected. If the CI covers r_0 , however, one may (a) conclude that the analyzed data set does not provide sufficient evidence to make a decision of rejection or no rejection of the tested null hypothesis and (b) recommend a replication study with a larger sample, in the hope that due to higher power then the CI will be sufficiently narrow to enable a decision as indicated above in this paragraph. When the null and alternative hypotheses have respectively the reverse directions to those in hypotheses (8) and (9), one proceeds conversely by complete analogy, using the same confidence level CI.

Last but not least, if the simple hypothesis (7) is to be tested against a one-tailed alternative, i.e., $r > r_0$ or $r < r_0$, which is of substantive relevance, then one can proceed as follows. First, impose in the fitted model (1) and (2) the nonlinear constraint

$$\tau_{00}/\sigma^2 = r_0 \quad (10)$$

and halve the resulting p -value associated with the pertinent likelihood ratio test (chi-square goodness of fit value) in this nested model (cf. Rabe-Hesketh & Skrondal, 2012). If the sample estimate of r is in the direction of the alternative and this halved p -value is lower than a preset significance level, for example, α , rejection of the H_0 in (7) is warranted; otherwise that null hypothesis H_0 is not rejected.

We demonstrate next on empirical data the preceding discussion in this note.

Illustration on Data

We employ for our aims in this section data from a study conducted by Mortimore, Sammons, Stoll, Lewis, and Ecob (1988; see du Toit & du Toit, 2001, also for the original data). To this end, we use scores on a Mathematics test that stem from $N = 1192$ students attending $G = 49$ schools randomly sampled from the London area. We begin by fitting the unconditional latent variable model (1) and (2) (cf. Raykov, 2011), whose fit is perfect since it is saturated: chi-square value (χ^2) = 0, degrees of freedom (df) = 0. (The needed *Mplus* source code is provided in Appendix A, with annotating comments following an exclamation mark within pertinent row.) The obtained estimates thereby of the within-group and between-group variances as well

as their ratio r are correspondingly as follows (standard errors are presented in parentheses, and hat denotes estimate):

$$\begin{aligned}\hat{\sigma}^2 &= 81.814 \text{ (3.415)} \\ \hat{\tau}_{00} &= 3.098 \text{ (1.330)}, \text{ and} \\ \hat{r} &= 0.038 \text{ (.016)}.\end{aligned}\tag{11}$$

Using next the R-function “ci.r” in Appendix B with these results for the variance ratio, r , its 95% CI is furnished as (.017, .087). This CI could be interpreted as a range of plausible values, at the 95% confidence level, for the number of times the between-school variance on this test is smaller than the within-school variance on it, with this range stretching from 2% to 9% (rounded off). That is, one could consider plausible (at this confidence level) that the between-school variance on the used test is approximately between 11 and 50 times smaller than the within-school variance. This result may be interpreted as suggesting that the between-school differences on the mathematics test under consideration represent a small fraction of the within-school differences in the student scores on it, and that schools are rather homogeneous on this test relative to student differences within school. Further substantive interpretation of this funding (confidence interval) would be left to the expert in the pertinent subject-matter field.

Conclusion

This article was concerned with a procedure for examining the degree to which between-cluster differences surpass or alternatively are dominated by within-cluster differences in two-level studies. A latent variable modeling method for point and interval estimation of the ratio of between to within group variance on a response variable was outlined. The discussed approach is useful in multilevel studies as a complement to the popular ICC and contributes to a more pragmatically oriented interpretation of the relationship between within- and between-cluster variances. As a possible addendum to point and interval estimates of the ICC (e.g., Raykov, 2011), the present procedure permits empirical researchers to reach also more informed conclusions about the extent to which the classical assumption of independence of units of analysis may be violated (see also Rabe-Hesketh & Skrondal, 2012).

The described method possesses several limitations that need to be pointed out. While it can be readily extended to three-level studies, for example, following the developments in Raykov et al. (2016), the procedure is presently available for (approximately) continuous normal response variables (see Raykov & Marcoulides, 2015, for a possible extension approach to the discrete case). While it could be argued that up to mild violations of normality could be handled via robust maximum likelihood (ML) estimation unless with piling of cases at a scale end (e.g., L. K. Muthén & Muthén, 2015), which may be potentially also applicable with items

having at least five to seven response options, additional research is needed to address this conjectured robustness. Furthermore, the underlying estimation procedure is best used with large samples both with respect to the number of clusters and observations within them (e.g., Raudenbush & Bryk, 2002; Raudenbush & Xiao-Feng, 2001). This stems from the fact that the method rests on ML estimation that itself is grounded in an asymptotic statistical theory. Future research is thus also encouraged that should be concerned with the development of possible guidelines that could be followed in evaluating sample size requirements of level-2 and level-1 units, at which one could rely in practice on the ML asymptotic theory.

In conclusion, this article provides educational and behavioral researchers with a readily applicable means of interval estimation for the ratio of between-group to within-group variance, which allows more pragmatically oriented interpretation of the fundamental variance decomposition in hierarchical designs and aids more informed decisions about the degree of interdependence of outcome scores on studied response variables. The outlined procedure may also be particularly beneficial as a complement to the popular ICC, in that it permits examining more thoroughly the clustering effect in multilevel data and could help in choosing if needed between traditional, single-level models and more complex, multilevel models for use in ensuing analyses.

Appendix A

Mplus Source Code for Evaluation of the Ratio of Between- to Within-Group Variance in Two-Level Studies

```

TITLE:      INTERVAL ESTIMATION OF THE RATIO OF BETWEEN
            TO WITHIN GROUP VARIANCE IN TWO-LEVEL STUDIES
DATA:      FILE = <NAME OF RAW DATA FILE>; ! Data from Mortimore et al. (1988).
VARIABLE:  NAMES = SCHOOLID STUDID GENDER IQ MATH1-MATH3 ENG1-ENG3 UNIT;
            USEVARIABLE = math1; ! Name of outcome variable Y.
            CLUSTER = SCHOOLID; ! Name of level-2 identifier.
ANALYSIS:  TYPE = TWOLEVEL; ! Requests two-level modeling.
            ESTIMATOR = ML; ! Requests use of ML estimation method, for ML estimate of  $r$ .
MODEL:
            %WITHIN% ! Defines level-1 model (within-group model).
            MATH1 (P1); ! Assigns parametric symbol to within-variance.
            %BETWEEN% ! Defines level-2 model (between-group model).
            MATH1 (P2); ! Assigns parametric symbol to between-cluster variance.
MODEL CONSTRAINT: ! See Equation (5).
            NEW (R); ! Introduces a 'place-holder' for the ratio  $r$  in Equation (5).
            R = P2/P1; ! Defines the ratio  $r$  as a 'new parameter'.
OUTPUT:    CINTERVAL; ! Requests standard errors for all model parameters.

```

Note 1. After a title, the name of the raw data file is to be stated and its columns assigned corresponding variable names. The outcome variable of interest is then selected, using the subcommand USEVARIABLE of the DATA command, and the level-2 unit identifier is given with

the subcommand CLUSTER. The ANALYSIS command requests two-level modeling with ML estimation, to render an ML estimate of the critical ratio r in Equation (5). The MODEL command first defines the within-group model (level-1 model; Raudenbush & Bryk, 2002), which is achieved via the keyword %WITHIN%, where only the residual variance is a parameter (see Equation 1). Then the between-group model definition follows (level-2 model), accomplished with the keyword %BETWEEN%, where the fixed effect γ_{00} is a parameter by default while the between group variance is declared as a parameter (see Equation 2). Thereby, the within-group variance σ^2 and the between group variance τ_{00} are assigned the parametric symbols P1 and P2, respectively. These symbols are used in the subsequent definition of these variances' ratio, r , in the MODEL CONSTRAINT section, where a "place holder" is first requested for this "new parameter" r (that is not a model parameter) via the NEW keyword, and then that ratio is formally defined (see Equation 5). An application of the delta method to produce an approximate standard error for this ratio is requested in the end in the OUTPUT command.

Note 2. Use of the R-function in Appendix B can subsequently be made to furnish a confidence interval for the ratio r , using its estimate and standard error obtained with the present *Mplus* command file.

Appendix B

R-Function for Interval Estimation of the Ratio in Equation (5) of Between- to Within-Group Variance in Two-Level Studies

```
ci.r = function(r, se) { # interval estimation of the ratio  $r$  in Equation (5) .
  l = log(r) # get the estimate of the ratio  $r$  from output of code in App. A.
  sel = se/r # get its S.E. from that output also.
  ci_l_lo = l - 1.96 * sel # routine CI construction for logged ratio  $r$ .
  ci_l_up = l + 1.96 * sel
  ci_lo = exp(ci_l_lo) # back-transforming for lower endpoint of  $r$ 's CI.
  ci_up = exp(ci_l_up) # back-transforming for upper endpoint of  $r$ 's CI.
  ci = c(ci_lo, ci_up)
  ci # print to screen the finally obtained CI for  $r$ .
}
```

Note. To furnish a confidence interval for the ratio r of between- to within-cluster variance at another confidence level, use in this file the appropriate cutoffs of the standard normal distribution when furnishing a confidence interval for its logarithm (e.g., 1.645 if a 90% CI of the ratio r is required per se).

Acknowledgments

We thank G. Mels for a valuable discussion on bounded parameter estimation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Since the null hypothesis $H_0: \tau_{00} = 0$, also referred to later, spans a lower-dimensional subspace in the three-dimensional parameter space of the mixed model (3), it is associated with probability (mass) 0 (in the latter space; e.g., Apostol, 2013). That is, in pragmatic terms, one can consider the ratio (Equation 5) as being positive (apart from an event with zero probability that need not be of relevance in an empirical setting; notice also the earlier assumption of positive within-group variance).
2. It is important to note that the CI-based approach that follows in the main text (regardless how the pertinent CI is obtained) cannot be used for testing this hypothesis of vanishing ratio r against the alternative of it being positive. The reason is that with probability 1 the point 0 will not be included in the CI (see also Note 1). In particular, if the bootstrap method is used alternatively for obtaining a CI, the resulting sampling distribution of the estimate of r , for any initial finite sample and number of resamples with replacement (at any size up to the original one), will contain with probability 0 the point 0. The reason is, as mentioned in Footnote 1, the fact that this null hypothesis (H_0) spans a lower-dimensional subspace in the parameter space of model (3), which subspace has mass (probability) of 0.

References

- Apostol, T. G. (2013). *Calculus*. New York, NY: Wiley.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, England: Cambridge University Press.
- du Toit, S. H. C., & du Toit, M. (2001). *Interactive LISREL 8: User's guide*. Lincolnwood, IL: Scientific Software International.
- Goldstein, H. (2011). *Multilevel statistical models*. London, UK: Arnold.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters, the junior years*. Wells, England: Open Books.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 87-117.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling with Stata*. College Station, TX: Stata Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6, 387-401.
- Raykov, T. (2010). Proportion of third-level variance in multilevel studies: A note on an interval estimation procedure. *British Journal of Mathematical and Statistical Psychology*, 63, 417-426.

- Raykov, T. (2011). Intra-class correlation coefficients in hierarchical designs: Evaluation using latent variable modeling. *Structural Equation Modeling*, 18, 73-90.
- Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parametric functions in covariance structure models. *Structural Equation Modeling*, 11, 659-675.
- Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. New York, NY: Taylor & Francis.
- Raykov, T., & Marcoulides, G. A. (2014). Identifying useful auxiliary variables for incomplete data analyses: A note on a group difference examination approach. *Educational and Psychological Measurement*, 74, 537-550.
- Raykov, T., & Marcoulides, G. A. (2015). Intraclass correlation coefficients in hierarchical design studies with discrete response variables: A note on a direct interval estimation procedure. *Educational and Psychological Measurement*, 75, 1063-1070.
- Raykov, T., Patelis, T., Marcoulides, G. A., & Lee, C.-L. (2016). Examining intermediate omitted levels in hierarchical designs via latent variable modeling. *Structural Equation Modeling*, 23, 111-115. doi:10.1080/10705511.2014.938186