DISCUSSION PAPER SERIES

IZA

IZA DP No. 3178

A Comment on Variance Decomposition and Nesting Effects in Two- and Three-Level Designs

Spyros Konstantopoulos

November 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Comment on Variance Decomposition and Nesting Effects in Two- and Three-Level Designs

Spyros Konstantopoulos
Northwestern University
and IZA

Discussion Paper No. 3178
November 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3178
November 2007

ABSTRACT

A Comment on Variance Decomposition and Nesting Effects in Two- and Three-Level Designs

Multilevel models are widely used in education and social science research. However, the effects of omitting levels of the hierarchy on the variance decomposition and the clustering effects have not been well documented. This paper discusses how omitting one level in three-level models affects the variance decomposition and clustering in the resulting two-level models. Specifically, I used the ANOVA framework and provided results for simple models that do not include predictors and assumed balanced nested data (or designs). The results are useful for teacher and school effects research as well as for power analysis during the designing stage of a study. The usefulness of the methods is demonstrated using data from Project STAR.

JEL Classification: C00

Keywords: variance decomposition, nested designs, clustering

Corresponding author:

Spyros Konstantopoulos
School of Education and Social Policy
Northwestern University
2120 Campus Drive
Evanston, IL 60208
USA
E-mail: spyros@northwestern.edu

Many populations of interest in education, psychology, and the social sciences have multilevel structure (e.g., students are nested within classrooms and classrooms are nested within schools, individuals are nested within neighborhoods, which are nested within cities). Because individuals within aggregate units (e.g., classrooms or schools) are often more alike than individuals in different units, this nested structure produces what is called in the sampling literature clustering effects (see, e.g., Kish, 1965). First, clustering effects need to be taken into account when analyzing data with nested structures. For example, one shortcoming of ignoring dependencies in the data is that the estimated standard errors of the regression coefficients are typically underestimated, leading to liberal tests of significance and an inflated probability of making a Type I error. In sampling methodology the clustering effects are captured by the design effect that is used to correct the standard errors of regression estimates (see Cochran, 1977; Kish, 1965; Lohr, 1999). In education and the social sciences the effects of clustering have been well addressed by multilevel models the last 20 years (Goldstein, 2003; Longford, 1993; Raudenbush & Bryk, 1986, 2002; Snijders & Bosker, 1999). Such models take into account the clustering effects in the estimation of the standard errors of the regression coefficients. Especially in two-level models with one level of clustering researchers have shown the degree of underestimation of the standard errors of the estimates that takes place when one ignores the dependency of the data at the second level (e.g., Raudenbush & Bryk, 1986, 2002).

Second, clustering effects need to be taken into account when designing studies with nested structure that do not follow a simple random sampling scheme. Multilevel or nested designs are easily understood by recognizing the sampling used at different levels of the population hierarchy. For example, the clustering effect in a two-level design that follows a two-stage cluster sampling (e.g., sample schools in the first stage, and then sample students within these schools at the second stage) is typically defined via an intraclass correlation (see Cochran, 1977; Lohr, 1999). This intraclass correlation is involved in computations of statistical power (an important aspect of study design) that are performed at the designing stage of a study (see Donner & Klar, 2000; Hedges & Hedberg, 2007, Murray, 1998; Raudenbush & Liu, 2000). In two-level designs with one level of clustering, researchers have documented the importance of including the intraclass correlation in power computations and have discussed the overestimation in statistical power that takes place when one ignores the effect of clustering at the second level (Hedges & Hedberg, 2007; Murray, 1998; Raudenbush & Liu, 2000; Snijders & Bosker, 1999).

Although two-level models are common practice in education and the social sciences, statistical analyses do not always involve two levels. For example, educational researchers have demonstrated the usefulness of applying three-level models to nested achievement data (see e.g., Bryk & Raudenbush, 1988; Nye, Konstantopoulos, & Hedges, 2004; Rowan, Correnti, & Miller, 2002). These studies have shown empirically that there are important clustering effects at the second and at the third level of the hierarchy. In addition, in study design clustering effects can take place at the second and at the third level of the hierarchy. Consider a three-level design that follows a three-stage cluster sampling (e.g., sample schools in the first stage, sample classrooms in the second stage, and then sample students within these classrooms at the third stage). The clustering effects are in this case typically defined via two intraclass correlations, one at the second and one at the third level (see Cochran, 1977; Lohr, 1999). These intraclass correlations are involved in computations of statistical power that are performed at the designing stage of three-level studies with two levels of nesting (see e.g., Konstantopoulos, in press).

When a three-level design follows a three-stage sampling scheme the total variance in the outcome is decomposed into three components: the within level-2 between level-1 unit (e.g., between students within classrooms) variance, σ_e^2; the within level-3 between level-2 unit (e.g., between classrooms within schools) variance, τ^2; and the between level-3 unit (e.g., between schools) variance, ω^2. Then, the total variance in the outcome is defined as

$$\sigma_T^2 = \sigma_e^2 + \tau^2 + \omega^2 . \tag{1}$$

In such three-level designs two intraclass correlations are needed to describe the variance component structure. These are defined as the second level intraclass correlation:

$$\rho_2 = \frac{\tau^2}{\sigma_T^2} \tag{2}$$

and the third level intraclass correlation

$$\rho_3 = \frac{\omega^2}{\sigma_T^2} \tag{3}$$

where the subscripts 2 and 3 indicate the level of the hierarchy.

However, it is not uncommon in practice to omit one level and treat data (or designs) with three sources of variation (e.g., within-classroom, between-classroom, and between-school) as data (or designs) with two levels of variation (e.g., within-school, between-school variation, or within-classroom, between-classroom variation). That is, frequently, analyses of nested data are conducted without including all levels of the hierarchy. Omitting a level of the hierarchy in analyses is sometimes a matter of convenience and other times a necessity. In education, for instance, classroom identifiers are not always available and thus the educational researcher in such cases may conduct analyses employing two-level models (where students are nested within schools). Similarly, in the designing phase of a study sometimes information about clustering effects (such as intraclass correlations) is not available for all levels of the hierarchy. In such cases, power analyses are conducted omitting one (or more) levels of the hierarchy due to lack of information.

When one source of variation (or clustering) is ignored in three-level models, either at the second or at the third level, the remaining two levels of variation have to absorb that variation, because the total variation in the outcome remains constant. However, the mechanism of the variance decomposition among two- and three-level data (or designs) is not that clear. In particular, consider data (or study designs) that follow a three-level nested structure with two levels of nesting. Suppose that the researcher treats the data (or the design) employing two-level models, by ignoring either the second or the third level. However, omitting a level will affect the estimates of the variance components and the clustering effects for the remaining levels. This paper uses ANOVA results for balanced designs and provides derivations about how the variance decomposition in three-level models is changed when only two-levels of the hierarchy are taken into consideration. I discuss the simplest multilevel model where no covariates are included at any level for two distinct cases. First I show how the variance decomposition takes place when the middle level is omitted, and then I show how the variance decomposition takes place when the third level is ignored.

A Two-Way Nested Random Model

Suppose that the data (or the design) follow indeed a three-level structure with two levels of nesting (at the second and third level). Consider a three-level unconditional model with no covariates at any level of the hierarchy. Within the ANOVA framework this is a two-way nested random model (e.g., students are nested within classrooms, and classrooms are nested within schools). The structural model equation for the l^{th} level-1 unit in the k^{th} level-2 unit in the j^{th} level-3 unit is

$$Y_{jkl} = \mu + \beta_{\text{j}} + \gamma_{\text{jk}} + \varepsilon_{jkl}, \tag{4}$$

where μ is the grand mean, β_j is the random effect of the level-3 unit j (j = 1,…, m), γ_{jk} is the random effect of level-2 unit k (k = 1,…, p) within level-3 unit j, and ε_{jkl} is the error term of the level-1 unit l (l = 1,…, n) within level-2 unit k, within level-3 unit j. The level-1, level-2, and level-3 random effects are normally distributed with a mean of zero and variances σ_e^2, τ^2, and ω^2 respectively. Following Kirk (1995) and Searle, Casella, and McCullogh (1992) I define the total sums of squares in this three-level model as

$$SS_T = SS_1 + SS_2 + SS_3 \tag{5}$$

where the subscripts *1*, *2*, and *3* indicate the level of the hierarchy. The expected value of the sums of squares at the first level is

$$E(SS_1) = mp(n-1)\sigma_e^2, \tag{6}$$

the expected value of the sums of squares at the second level is

$$E(SS_2) = m(p-1)(\sigma_e^2 + n\tau^2), \tag{7}$$

and the expected values of the sums of squares at the third level is

$$E(SS_3) = (m-1)(\sigma_e^2 + n\tau^2 + pn\omega^2), \tag{8}$$

assuming m level-3 units, p level-2 units, and n level-1 units.

Case A: Omitting the Second Level of the Hierarchy

First, suppose that the second level (e.g., classroom) of the three-level structure is omitted, and the model is reduced to two-levels (e.g., student and schools). Then, the sums of squares at the second level (e.g., school) and at the first level (e.g., student) of thr resulting two-level model are defined as

$$\widetilde{SS}_2 = SS_3, \tag{9}$$

and

$$\widetilde{SS}_1 = SS_1 + SS_2 \tag{10}$$

respectively. The objective is to compute the expected values of the first and second level variances $\tilde{\sigma}_1^2$, $\tilde{\omega}_2^2$. Specifically, using equations 7, 8, and 10 the expected value of the first level variance $\tilde{\sigma}_1^2$ is

$$E(\tilde{\sigma}_1^2) = \frac{E(SS_1) + E(SS_2)}{m(pn-1)} = \frac{m(p-1)(\sigma_e^2 + n\tau^2) + mp(n-1)\sigma_e^2}{m(pn-1)} = \sigma_e^2 + \left(\frac{n(p-1)}{pn-1}\right)\tau^2 . \tag{11}$$

The above equation indicates that when the middle level (in a three-level structure) is omitted the first level variance in the resulting two-level model is the sum of the first level variance and a portion of the second level variance in the three-level model. Notice that when *n* (e.g., the number of students within each classroom) becomes infinitely large the term $\left(\frac{n(p-1)}{pn-1}\right)\tau^2 \to 0$ tends to zero, and when *p* (e.g., the number of classrooms per school) becomes infinitely large the term $\left(\frac{n(p-1)}{pn-1}\right)\tau^2 \to \tau^2$ tends to τ^2. This suggests that when the number of level-1 units is large and the number of level-2 units is small (in a three-level structure) the middle level variance does not affect much the first level variance of the resulting two-level model.

Similarly, since

$$E(\tilde{\omega}_2^2) = \frac{E(\widetilde{MS}_2) - E(\widetilde{MS}_1)}{pn} \tag{12}$$

and

$$E(\widetilde{MS}_2) = E(MS_3) = \frac{E(SS_3)}{(m-1)},\ E(\widetilde{MS}_1) = \frac{E(\widetilde{SS}_1)}{m(pn-1)} \tag{13}$$

and using equations 8 and 11 the expected value of $\tilde{\omega}_2^2$ is

$$E(\tilde{\omega}_2^2) = \frac{E(SS_3)/(m-1) - E(\tilde{\sigma}_1^2)}{pn} = \omega^2 + \left(\frac{n-1}{pn-1}\right)\tau^2. \tag{14}$$

The above equation indicates that when the middle level (in a three-level structure) is omitted the second level variance of the resulting two-level model is the sum of the third level variance and a portion of the second level variance in the three-level model. Notice that when n (e.g., the number of students within each classroom) becomes infinitely large the term $\left(\frac{n-1}{pn-1}\right)\tau^2 \to 1$ tends to one, and when p (e.g., the number of classrooms per school becomes infinitely large the term $\left(\frac{n-1}{pn-1}\right)\tau^2 \to 0$ tends to zero. This suggests that when the number of the middle-level units (in a three level structure) is large the middle-level variance τ^2 does not affect much the second level variance of the resulting two-level model. However, when the number of level-1 units (in a three-level structure) is large the second level variance of the resulting two-level model is the sum of the second and the third level variances in the three-level model. Notice that the sum of equations 11 and 14 is

$$E(\tilde{\sigma}_{ws}^2) + E(\tilde{\omega}_{bs}^2) = \sigma_e^2 + \left(\frac{n(p-1)}{pn-1}\right)\tau^2 + \omega^2 + \left(\frac{n-1}{pn-1}\right)\tau^2 = \sigma_e^2 + \tau^2 + \omega^2.$$

It is straightforward to derive the clustering effect in this case. Suppose that the nesting effect is expressed via an intraclass correlation $\tilde{\rho}_2$. Then, using equation 11 and 14 it follows that

$$\tilde{\rho}_2 = \frac{\tilde{\omega}_2^2}{\tilde{\sigma}_1^2 + \tilde{\omega}_2^2} = \frac{\omega^2 + \left(\frac{n-1}{pn-1}\right)\tau^2}{\sigma_e^2 + \tau^2 + \omega^2} = \rho_3 + \left(\frac{n-1}{pn-1}\right)\rho_2 , \tag{15}$$

which indicates that when the number of level-1 units n (in a three-level structure) is quite large the intraclass correlation in the resulting two-level model is the sum of the intraclass correlations at the second and at the third level in the three-level model. However, when the number of the middle-level units p (e.g., classrooms) is quite large the intraclass correlation in the resulting two-level model is simply the intraclass correlation at the third level in the three-level model.

Case B: Omitting the Third Level of the Hierarchy

Second, suppose that the third level (e.g., school) in the three-level structure is omitted, and the model is reduced to two-levels (e.g., students and classrooms). Then, the sums of squares at the second level (e.g., classroom) and at the first level (e.g., student) of the resulting two-level model are defined as

$$\widehat{SS}_2 = SS_2 + SS_3 , \tag{16}$$

and

$$\widehat{SS}_1 = SS_1 \tag{17}$$

respectively. The objective is to compute the expected values of the first and second level variances $\widehat{\sigma}_1^2$, $\tilde{\tau}_2^2$. The expected value of the first level variance in the resulting two-level model is simply the first level variance in the three-level model, namely

$$E(\widehat{\sigma}_1^2) = \sigma_e^2 . \tag{18}$$

Similarly, since

$$E(\tilde{\tau}_2^2) = \frac{E(\widehat{MS}_2) - E(\widehat{MS}_1)}{n} = \frac{E(\widehat{SS}_2)/(mp-1) - E(\widehat{\sigma}_1^2)}{n} \quad (19)$$

and

$$E(\widehat{SS}_2) = E(SS_2) + E(SS_3)\,, \quad (20)$$

and using equations 6, 7, and 18, and 20 the expected value of $\tilde{\tau}_2^2$ is

$$E(\tilde{\tau}_2^2) = \frac{\left\{ m(p-1)(\sigma_e^2 + n\tau^2) + (m-1)(\sigma_e^2 + n\tau^2 + pn\omega^2) \right\}/(mp-1)}{n} - \frac{\sigma_e^2}{n}$$

which reduces to

$$E(\tilde{\tau}_2^2) = \tau^2 + \left(1 - \frac{p-1}{mp-1}\right)\omega^2 . \quad (21)$$

The above equation indicates that when the third level (in a three-level structure) is omitted, the second level variance in the resulting two-level model is the sum of the second level variance and a portion of the third level variance in the three-level model. Notice that when p (e.g., the number of classrooms within each school) becomes infinitely large the term $\left(1 - \frac{p-1}{mp-1}\right)\omega^2 \to 0$ tends to zero, and when m (e.g., the number of schools) becomes infinitely large the term $\left(1 - \frac{p-1}{mp-1}\right)\omega^2 \to \omega^2$ tends to ω^2. This suggests that when the number of the middle-level units (in a three-level structure) is large the third level variance does not affect much the second level variance in the resulting two-level model. However, when the number of level-3 units (in a three-level structure) is large the second level variance in the resulting two-level model is the sum of the second and third level variances in the three-level model. Also, notice that in this case the first level variance in the three-level model and the first level variance in the resulting two-level model are the same. As with the previous case A, suppose that the clustering effect is expressed via an intraclass correlation $\widehat{\rho}_2$. Then, using equations 18 and 21 it follows that

$$\widehat{\rho}_2 = \frac{\tilde{\tau}_2^2}{\widehat{\sigma}_1^2 + \tilde{\tau}_2^2} = \frac{\tau^2 + \left(1 - \frac{p-1}{mp-1}\right)\omega^2}{\sigma_e^2 + \tau^2 + \left(1 - \frac{p-1}{mp-1}\right)\omega^2}, \qquad (22)$$

which indicates that when the number of level-3 units is quite large the intraclass correlation in the resulting two-level model is the sum of the intraclass correlations at the second and third levels in the three-level model.

Example

To show the usefulness of the methods presented in this paper I used data from Project STAR (see Nye, Hedges, & Konstantopoulos 2000). Specifically I ran a three-level unconditional model (no covariates at any level) using kindergarten data from project STAR to model mathematics achievement. I standardized the outcome so that its total variance is one. There were about 18 students per classroom (n = 18), four classrooms per school (p = 4) and 79 schools (m = 79). The results of the three-level analysis indicated that the variances at the first, second, and third levels were respectively 0.709, 0.126, 0.165.

First, suppose that the middle level (e.g., classrooms) is omitted. Then, using equations 11 and 14 I computed the first and second level variance of the resulting two-level model (e.g., students nested within schools) as

$$\tilde{\sigma}_1^2 = 0.709 + \left(\frac{18(4-1)}{72-1}\right)0.126 = 0.805,$$

and

$$\tilde{\omega}_2^2 = 0.165 + \left(\frac{18-1}{72-1}\right)0.126 = 0.195.$$

The two-level HLM analyses of the data provided almost identical variance components estimates: 0.80 for the first level variance, and 0.20 for the second level variance. Because the total variance is one in this example, the clustering effect expressed as an intraclass correlation is 0.195 as well, which is about 2/3 of the sum of the clustering effects (= 0.126 + 0.165 = 0.291) in the three-level model. Specifically,

$$\tilde{\rho}_2 = 0.165 + \left(\frac{18-1}{72-1}\right)0.126 = 0.195.$$

Second, suppose that the third level (e.g., schools) is omitted. Then, using equations 18 and 21 I computed the first and second level variance of the resulting two-level model (e.g., students nested within classrooms) as

$$\tilde{\sigma}_1^2 = 0.709,$$

and

$$\tilde{\tau}_2^2 = 0.126 + \left(1 - \frac{4-1}{316-1}\right)0.165 = 0.289.$$

The two-level HLM analyses of the data provided almost identical variance components estimates: 0.71 for the first level variance, and 0.29 for the second level. A small discrepancy in the estimates is expected since the level-1, level-2, and level-3 units were computed approximately (assuming a balanced design). Because the total variance is one in this example, the clustering effect expressed as an intraclass correlation is 0.289 as well, which is about the sum of the clustering effects (= 0.126 + 0.165 = 0.291) in the three-level model. This is expected since the number of level-3 units was large in this case ($m = 79$). Specifically,

$$\widehat{\rho}_2 = \frac{0.126 + \left(1 - \frac{4-1}{316-1}\right)0.165}{0.709 + 0.126 + \left(1 - \frac{4-1}{316-1}\right)0.165} = 0.289 .$$

In sum, this paper showed that omitting a level in three-level models affects the variance decomposition and clustering effects in the resulting two-level models. The results are presented in algebraic expressions that are easy to use. These results are useful for education and social science researchers (e.g., in teacher and school effects research) since they indicate what part of the middle level (e.g., classroom) variance (in a three-level model) is distributed to the first (e.g., within school) and the second (e.g., between school) level (in the resulting two-level model), or what part of the third level (e.g., school) variance (in a three-level model) is included in the second level (e.g., between classroom) variance (in the resulting two-level model). These results are also useful for power computations during the designing stages of a study, since they provide a guide about how the clustering effects change from three- to two-level designs.

References

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects. A three-level hierarchical linear model. *American Journal of Education*, *97*, 65-108.

Cochran, W. G. (1977). *Sampling techniques*. New York: Wiley.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Hedges, L. V., & Hedberg, E. (2007). Intraclass correlation values for planning group randomized trials in Education. *Educational Evaluation and Policy Analysis, 29,* 60-87.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing.

Konstantopoulos, S. (in press). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness.*

Lohr, S. L. (1999). *Sampling: Design and analysis.* Duxbury Press.

Longford, N T. (1993). *Random coefficient models.* New York: Oxford University Press.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Nye, B, Hedges, V. E., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, *37*, 123-151.

Nye, B, Konstantopoulos, S., & Hedges, V. E. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237-257.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1-17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trails. *Psychological Methods*, *5*, 199-213.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record*, *104*, 1525-1567.

Searle, S. R., Casella, G., & McCuloch, C. E. (1992). *Variance components.* New York: Wiley.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London: Sage.