The Consequence of Ignoring a Level of Nesting in Multilevel Analysis: A Comment

Georges Van Landeghem, Bieke De Fraine, and Jan Van Damme
Centre for Educational Effectiveness and Evaluation
K. U. Leuven

This short contribution is a comment on M. Moerbeek’s exploration of consequences of ignoring a level of clustering in a multilevel model, which was published in the first issue of the 2004 volume of *Multivariate Behavioral Research*. After having recapitulated the framework and extended the results of Moerbeek’s study, we formulate two critical notes. First, we point at the incompleteness of the conclusions drawn by Moerbeek from the analytical work. The second note is concerned with the limitations of the framework itself.

RECAPITULATION

Framework of the Analytical Study

Moerbeek’s (2004) study is centered around the multilevel models (Goldstein, 1995; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) that simultaneously satisfy the following six conditions:

1. The model is completely hierarchical, that is, the basic level is nested within the second level of clustering, the second level is nested within the third, and so on, up to the top level. Cross-classified structures (Goldstein & Sammons, 1997; Rasbash & Goldstein, 1994)—which occur, for example, when the clustering of pupils in classes of consecutive grades is taken into account (Van Landeghem, Van Damme, Opdenakker, De Fraine, & Onghena, 2002)—are excluded.

Part of this research was supported by the Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs, through the Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt. Correspondence concerning this article should be addressed to G. Van Landeghem, K. U. Leuven, Centrum voor Onderwijseffectiviteit en -Evaluatie, Dekenstraat 2, B-3000 Leuven, Belgium. E-mail: Georges.VanLandeghem@ped.kuleuven.ac.be
2. The model is completely balanced. This means that each level of clustering in the model is balanced, that is, all its groups contain the same number of basic level units. This is an important restriction, as many data sets with a multilevel structure are unbalanced. In educational research, for example, all pupils of a number of classes with different sizes may be included in the data set (e.g., Burns & Mason, 2002).

3. Only the intercept is allowed to vary randomly. The coefficients of the explanatory variables must be fixed. Random slopes—expressing, for example, that the effect of a school characteristic on a classroom-level outcome can vary across states (Swanson & Stevenson, 2002)—are not allowed.

4. The response variable and the latent variables (residuals) at all levels have a normal distribution. Thus, multilevel models for a dichotomous response, counts, etcetera, (see, e.g., Snijders & Bosker, 1999, chap. 14) are not considered.

5. Apart from explanatory variables associated with the top level—that is, variables that do not vary within any of the top-level clusters—every explanatory variable must be centered in a specific way. An explanatory variable associated with the $k^{th}$ level—which means that it does not vary within any of the clusters of the $k^{th}$ level and that it does vary (minimally perhaps) between those clusters—must be group mean centered (Raudenbush & Bryk, 2002) at the $(k + 1)^{th}$ level. In a three-level model of pupils within classes within schools, for example, this means that every pupil level explanatory variable must be centered within each class. Also, every class level variable must have a zero mean value within each school. The requirement holds for all explanatory variables, in particular also for those that are intrinsically dichotomous, such as gender. In order to satisfy the condition, the coding of a variable such as gender will need to become dependent on the class. In a class of 10 girls and 10 boys, the value 0.5 may be used to indicate a girl and −0.5 must then be used to indicate a boy; in a class with 5 girls and 15 boys, a class mean of zero may be obtained by coding 0.75 for a girl and −0.25 for a boy.

6. The estimators of the model parameters are defined by the maximum likelihood principle (also called the full information maximum likelihood principle and designated as ML, MLF, or FIML). This condition excludes the analyses where another frequently employed estimation principle, the restricted (or “residual”) maximum likelihood principle (REML, or MLR), is applied. (For more information concerning those estimation principles, see, among others, Goldstein, 1995; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999.)

In the next subsection, we discuss briefly the solution of a multilevel model satisfying the six conditions listed above. Before we do so, we need to comment on the fifth and sixth condition.

It is important to note how the fifth condition, concerning the centering of the explanatory variables, can be fulfilled in practice. Consider, for example, a sample
of $n_3$ schools, with $n_2$ classes sampled within each school, and $n_1$ pupils per class. Let us assume that a $n$ by 1 data column ($n = n_1 n_2 n_3$) of measurements of the socio-economic status (SES) of the pupils is available. After having calculated the $n_3$ school means and the $n_2 n_3$ class means of the SES, one can write each pupil’s SES as a sum of three components, namely, the mean SES of the pupil’s school, the difference between the mean SES of the pupil’s class and the mean SES of his/her school, and the difference between the pupil’s SES and the mean SES of his/her class. Provided that it is different from the zero vector, each one of these components constitutes an explanatory variable—a school, class, or pupil level variable, respectively—that satisfies the fifth condition. Thus, any explanatory variable can (at least technically) be made to satisfy the centering requirements mentioned above, provided that one is willing to allow a distinct coefficient for each one of its non-zero components (as recommended by Neuhaus & Kalbfleisch, 1998).

Regarding the sixth condition, it must be noted that it does not refer to a specific algorithm (EM, Newton-Raphson, IGLS, ..., see, among others, Snijders & Bosker, 1999, and references therein) to solve the FIML problem, and that such a specification is not necessary to obtain the results of the next section. (This is in contrast with Moerbeek’s reliance on the IGLS algorithm in a part of her analytical work.)

Summary of Analytical Results

Within Moerbeek’s (2004) framework, defined by these six conditions, it is possible to work out analytically how the point estimates of the fixed part coefficients, the point estimates of the variance components, and the (co)variances of the fixed effects estimators change when an intermediate level or the top level is ignored in the random part of the multilevel model, whilst leaving the explanatory variables in place and unchanged. Closed-form expressions of these changes can be obtained for models with any number of levels and any number of explanatory variables. The possibility of deriving these formulas rests on the highly regular structure of the covariance matrix of the response vector within this restrictive framework and on the particular relationship of this matrix with the explanatory variable data vectors centered according to the fifth condition. This special structure is usually destroyed by a nonhierarchical relationship between levels of clustering, the presence of random slopes, unbalancedness, or a different centering of the explanatory variables. (See Van Landeghem, Onghena, & Van Damme, 2001, who explained the breakdown of properties of orthogonal explanatory variables in the transition from ordinary regression to multilevel regression, and the rare conditions under which the property of separability is maintained in multilevel models.)

The analysis shows that parameter estimates of the multilevel model can be derived from the parameter estimates of a series of ordinary regression problems, one for each level. A detailed description of the relationship between these ordinary re-
gression problems and the multilevel model, of the general formulas expressing
the changes when a level is ignored, and of their derivation, falls outside the scope
of this comment. It is available upon request (from the corresponding author). Here
we confine ourselves to listing some general conclusions. First, the point estimates
of the fixed part coefficients remain unchanged (as illustrated in Moerbeek’s,
2004, text on p. 134 in a numerical example with three levels of clustering). Sec-
ond, when an intermediate level of clustering is ignored, its variance component is
split in two (generally unequal) parts and added to the components of the adjacent
levels (see Moerbeek’s formula 14 on p. 137, for the case of a three-level model).
When the top level is ignored, its variance component is added to the component of
the new top level (as found by Moerbeek in the three-level case, see formula 15 on
p. 137). Third, when the \( k \)th level of clustering is ignored, only the (co)variances of
the fixed effects estimators of the explanatory variables associated with the \( k \)th and
the \( (k − 1) \)th level are affected. Fourth, the general analysis shows that—within
Moerbeek’s framework—the qualitative consequences of ignoring an intermediate
or the top level in a model with more than three levels are identical to those of ig-
noring the second level or the third (i.e., top) level in a three-level model.

For further reference within this short note it is sufficient to highlight the formu-
las for the changes in coefficient estimator variances in the case treated by
Moerbeek (2004), namely a three-level model with a single explanatory variable
(and an intercept). Following Moerbeek, we again talk about a sample of \( n_3 \)
schools, within each of which \( n_2 \) classes have been sampled, each class being rep-
resented by \( n_1 \) pupils.

When the single predictor is a pupil level variable and the class level is ignored,
the term

\[
\frac{n_1n_2 - n_1}{n_1n_2 - 1} \hat{\sigma}_2^2 \frac{1}{n_1n_2n_3s_X^2},
\]

is added to the variance of the slope estimator (Moerbeek, 2004, p. 139 and Table
3). Here \( \hat{\sigma}_2^2 \) is the class level variance component estimate in the complete
three-level model and \( s_X^2 \) is the variance of the \( n_1n_2n_3 \) by 1 predictor vector.

When the predictor is a class level variable and the class level is ignored, the term

\[
\frac{n_1n_2 - n_2}{n_1n_2 - 1} \hat{\sigma}_2^2 \frac{1}{n_2n_3s_X^2},
\]

is subtracted from the variance of the slope estimator. Here \( s_X^2 \) is the variance of
the \( n_2n_3 \) by 1 predictor data column.
When the predictor is a class level variable and the school level is ignored, the term

$$\hat{\sigma}_3^2 \frac{1}{n_2 n_3 s_X^2},$$

(3)

is added to the variance of the slope estimator (Moerbeek, 2004, p. 142 and Table 3). (Here $\hat{\sigma}_3^2$ refers to the school level.) Furthermore, the term

$$\frac{n_2 - 1}{n_2} \hat{\sigma}_3^2 \frac{1}{n_3},$$

(4)

is subtracted from the variance of the intercept estimator.

As a final illustration, we add the case of a grand mean centered school level predictor. (The more general case without the grand mean centering requirement is covered by the general analysis, which is available upon request.) When the school level is ignored, the term

$$\frac{n_2 - 1}{n_2} \hat{\sigma}_3^2 \frac{1}{n_3 s_X^2},$$

(5)

is subtracted from the variance of the slope estimator. (Here $s_X^2$ is the variance of the $n_3$ by 1 predictor data column.) Simultaneously, the term shown by Equation 4 is subtracted from the variance of the intercept estimator.

**COMPLETION OF THE ANALYSIS AND ADJUSTMENT OF THE CONCLUSIONS**

Problematic Reason for Excluding Explanatory Variables
Associated With the Ignored Level

Moerbeek (2004) did not consider the effect of ignoring a level of clustering on a predictor variable associated with the ignored level. This limitation was founded on the argument that, when a predictor variable associated with a given level (as defined by the fifth condition above) is present in the model, “it is realistic to assume that all identifiers at the latter level are well registered and available in the data set” (p. 134).

This line of reasoning, however, is not consistent with the rest of Moerbeek’s (2004) study. Moerbeek reported the effect of ignoring the class level on the
slope estimator of a *class-mean centered* pupil level variable (p. 139) and also the effect of ignoring the school level on the slope estimator of a *school-mean centered* class level variable (p. 142). But it is not common for raw measurements to have been group mean centered automatically by some implicit feature of the measurement procedure. The group mean centering transformation must usually be executed explicitly by the researcher, using the group membership data. How can this be done when the reason for ignoring a level of clustering is the absence of these group membership data? Moerbeek argued that when the identifiers of the $k^{\text{th}}$ level of clustering are missing, explanatory variables associated with the $k^{\text{th}}$ level can not be present. But it seems equally plausible to argue that explanatory variables associated with the $(k-1)^{\text{th}}$ level, with the required centering at the $k^{\text{th}}$ level, cannot be present.

With regard to the covariances of the slope estimators, the analysis within the present framework becomes rather uninspiring when explanatory variables from both levels are left out of the discussion, as they are the only ones that can be affected. We have, however, reasons to believe that, within Moerbeek’s (2004) analytical framework, the effect of ignoring the $k^{\text{th}}$ level of clustering on explanatory variables associated with the $(k-1)^{\text{th}}$ level as well as on explanatory variables associated with the $k^{\text{th}}$ level must be taken into account.

**Additional Reasons for Ignoring a Level of Clustering**

First, we would like to emphasize that there are potential reasons other than the lack of membership data for ignoring certain levels of clustering. Although direct software limitations on the number of levels are a thing of the past, the technical possibility of including numerous levels of clustering does not guarantee convergence. Even when the software does not hinder the setup of a complex random part directly—as it still may in the case of intricate nonhierarchical relationships between the modes of clustering—the iterative algorithm may fail to find the solution of the likelihood problem.

D’Agostino (2000), for example, intended to use a measurement within pupil within class within school model, with, among other explanatory variables, instructional (classroom) practice variables (derived from a questionnaire filled out by the teacher of the class). But the difficulty of modeling a class level in a longitudinal study forced D’Agostino to abandon the class level (the availability of the identifiers notwithstanding; p. 217). The study proceeded with the construction of measurement within pupil within school models, including also the instructional explanatory variables (which then, technically, are inserted at the pupil level; see D’Agostino, 2000, pp. 217–218).

**Known Variables at an Unknown Level**

Second, in practice there are data sets that contain level-$k$ variables without a registration of the level-$k$ clusters. This may occur, for example, when in a sample of
classes, the mathematics teacher of each class is asked to answer questions about his/her years of experience (Swanson & Stevenson, 2002, p. 7). When teacher identifiers are not registered, it is no longer possible to include a teacher level between the class and the school level in the random part. But the variable “teacher’s years of experience” may still be stored in the data set as a class variable and inserted in, say (as in Swanson & Stevenson, 2002), a class within school within state model. The years of experience of the teacher may have been stored as a class variable, but it still is a variable at the teacher level—classes belonging to the same teacher will have the same value—a level that is ignored in the multilevel model because its identifiers have not been registered.

Experimental Studies

Although the previous examples refer to observational data, multilevel modeling and related issues such as the effect of ignoring a level of clustering can also be relevant in experimental studies. Intervention studies with random assignment of the treatment condition occurring at a level of clustering, provide a special example. The consequences of using ordinary regression (ignoring the level of clustering involved in the randomization) instead of multilevel regression for the estimation of the intervention effect, were examined by Carvajal, Baumler, Harrist, and Parcel (2001) and by Moerbeek, van Breukelen, and Berger (2003). Obviously, here a lack of membership data for the clusters involved in the randomization is an unlikely reason for using ordinary regression. Moerbeek et al. suggested that analysts may resort to ordinary regression simply because multilevel regression is “relatively new and rather complex” (p. 347).

Note that, in general, experimental data may involve additional levels of clustering (not necessarily involved in the treatment randomization) and additional explanatory variables (other than the treatment condition), thus giving rise to similar issues as discussed above by means of examples of observational studies.

Correct Conclusion From the Analytical Work

In practice, reasons for ignoring levels of clustering that do not prohibit the inclusion of explanatory variables associated with the ignored level may prevail. Even the absence of the identifiers of a given level of clustering does not necessarily imply the absence of variables associated with that level of clustering. Therefore, a systematic overview of the consequences of ignoring a level of clustering within the present framework must also include the results such as Equations 2, 4, and 5, which show a decrease of estimator variance. It changes one of the main conclusions of Moerbeek’s (2004) study, namely that by ignoring a level of clustering the “standard errors of regression coefficients estimators may be overestimated” (p. 129, see also pp. 139, 142, 144, 147), into the standard errors of regression coeffi-
Explanatory Variables With Significant Components at Several Levels

Instances of explanatory variables with important components at and/or below a level of clustering which is ignored, are frequently encountered in multilevel studies. Ai (2002), for example, constructed three-level models of measurements within pupils within schools. The variable “math teacher encouragement” is measured at the basic level (i.e., as a time-varying pupil characteristic). In the model it is used both as an aggregated school level variable (in interaction with a basic level “math attitude” variable) and as a basic level variable (see Ai, 2002, Table 4, p. 14). It is quite plausible that this “math teacher encouragement” measure varies at the teacher level, a potential level of clustering that is ignored in the model. Similarly, Griffith (2002) asked pupils about “classroom instrumental support” (based on questions about the teachers; see Table 2 on p. 356) and “classroom expressive support” (based on questions about the class and classmates; see Table 2 on p. 356). We expect these variables to have a significant class level variance component (in a three-level decomposition). Griffith uses them as explanatory variables in a two-level model (pupils within schools) which ignores the class level (Griffith, 2002, Table 6 on p. 361). Konu, Lintonen, and Autio (2002) included a pupil level variable “good work atmosphere in class” (p. 192), which is likely to have an important component at the class level, in a two-level model of pupils within schools. Van den Broeck, Van Damme, and Opdenakker (2004) used class characteristics measured by means of a teacher questionnaire in a model that ignores the clustering at the teacher level. Such cases make up a third reason for including the results of Equations 2, 4, and 5 (and their general counterparts) in the discussion. In view of our analytical results, we expect the magnitude and direction of change in the estimator variance of a single coefficient enforced on several components of an explanatory variable to be dependent on details of the model and the data, and to be difficult to predict.

LIMITATION OF THE ANALYTICAL RESULTS

Thus, having completed Moerbeek’s (2004) analytical work, we have learned that the (co)variance of slope estimators may be estimated incorrectly when a level of clustering is ignored. As the main result of an analytical study, this is rather meagre. After all, a few well-chosen numerical examples may yield the same information (see Opdenakker & Van Damme, 2000). Is it not possible to draw more specific conclusions?
Within the framework of the six conditions listed above, we can be much more specific. When the $k^{th}$ level of clustering is ignored, the (co)variance of slope estimators referring to the $k^{th}$ level is always underestimated—as illustrated by Formulas 2 and 5—whereas the (co)variance of slope estimators referring to the $(k - 1)^{th}$ level is always overestimated—as in Formulas 1 and 3. Moreover the general counterparts of the Formulas 1 to 5 unambiguously spell out the dependency of those changes in the estimated covariance matrix on the cluster sizes, variance component estimates and data matrices. But is it possible to generalize outside this framework?

There is a strong demand for “rules of thumb” to guide multilevel analyses. This is quite understandable because analysts pay a double price for the transition from ordinary regression to multilevel regression. On one hand, multilevel modeling involves more steps and choices—which levels of clustering to consider, where to allow random slopes, how to choose between different centering options, etcetera. On the other hand, there are fewer rules available to guide the process—no straightforward concept of explained variance (Snijders & Bosker, 1999, chap. 7), no direct advantage of mutually orthogonal explanatory variables (Van Landeghem et al., 2001), etcetera. For analysts in this less comfortable modeling environment, rigorous analytical results tend to have the appeal of a life buoy. As we are convinced that overstretched rules of thumb are worse than no guidance at all, we want to emphasize the limited generalizability of the analytical results obtained within Moerbeek’s framework.

We question whether it is possible to “give a systematic overview of the consequence of ignoring a level of nesting in multilevel analysis” (Moerbeek, 2004, p. 131) based on such a special subset of multilevel models as the one defined by the six restrictions of Moerbeek’s framework. The exceptional structure of these models marks them out as uncommon and not representative of multilevel models in general, with, for example the property that none of the usual iterative multilevel software (MLwiN, HLM, etc., for references, see Snijders & Bosker, 1999, chap. 15) is needed to estimate their parameters: software for ordinary regression is sufficient. Moreover, a cursory look at the literature is enough to bring out the singularity of Moerbeek’s analytical framework. Examples of multilevel models that fall outside this framework abound: They are not completely balanced, have random slopes, are not completely hierarchical, enforce equal coefficients on different components of an explanatory variable, and so on.

One might argue that a multilevel model that does not satisfy the six conditions of Moerbeek’s (2004) subset may well be sufficiently close to a model within this subset to warrant an approximate solution, by solving the latter model (by means of ordinary regressions) and adopting this solution as an estimate of the parameters of the former model. We are, however, not aware of any examples of such an approach in practice. Why not? Probably because there is no reliable rule to know when a model is “sufficiently close” to a model within the subset, short of calculat-
ing the exact solution in the first place and comparing with one or more likely candidates within the subset. This requires more effort than applying the usual multilevel software.

Moerbeek’s (2004) attempt to widen the applicability of the analytically derived formulas outside the framework defined by the six conditions, by suggesting that mean values of cluster sizes may be used in the formulas in the case of unbalanced models (p. 145), is bound to run into a similar difficulty. Even apart from the fact that this suggestion only treats models that violate a single condition, namely the one of complete balancedness, there is no rule that tells us how far the unbalancedness may go before the (approximate) formulas are no longer useful. There is no reason to expect that a simple rule describing these limits exists. And even if it is there, waiting to be discovered, it may well tell us that the usefulness of the analytical formulas is very limited outside the subset of models for which they are exact.

What constructive advice do we propose to give, then, to analysts who knowingly ignore a level of clustering and worry about the influence on the estimated standard errors of the coefficients of the fixed part of the model? Our advice is that they proceed with their analyses as if the problem was not there, report and interpret in the usual way, and finally qualify the whole by clearly stating that they suspect that a potentially important level of clustering has been ignored. Unless their model satisfies the conditions of Moerbeek’s (2004) analytical framework—which is unlikely—we advise against speculating about the direction, let alone the size, of the shifts in estimator covariances (and standard errors and power) as a consequence of leaving out the level in question. Instead, we encourage them to direct their energy towards obtaining the resources and data that will enable them to include the ignored level in future analyses.

AFTERWORD

An analyst’s decision to include or not to include a particular level of clustering in a model involves a diversity of considerations. Elements of the rationale behind the development of multilevel models (such as the unit of analysis problem, a need to generalize to a wider population of clusters, design effects due to clustering, the incorporation of cross-level interactions, etc.) may play a role. Especially, the researcher must work towards consistency of the model with the theoretical viewpoint which he/she wishes to adopt, his/her knowledge and assumptions about the data generating process, and the specific research questions of the study. Moerbeek’s (2004) analysis refers to a phase after this complex decision. It confronts the problem encountered by a researcher who is convinced that a certain level of clustering should be incorporated, but is somehow prevented from doing so.
We have argued that, from a practical point of view, Moerbeek’s (2004) analytical work does not yield more information than what can be gathered from numerical examples. The prospective practical return of a simulation study, the third way to tackle the problem, seems limited. Simulation is most appropriate to bridge the gaps between several subsets of cases that yield to analytical treatment, thus covering a substantial range of cases found in practice. This requirement is not fulfilled for the problem at hand, as there is only a single analytically manageable point of reference, defined by the six conditions of Moerbeek’s framework.

With the knowledge available at this moment, it seems more worthwhile to put effort into the removal of obstacles preventing the inclusion of a relevant level of clustering than to speculate about the consequences of excluding the level from the analysis.

REFERENCES


*Accepted January 2005*