The interaction between groups and individuals: The challenge of statistically analysing cooperative learning

Ulrike Cress, Knowledge Media Research Center, 72072 Tübingen,
Konrad-Adenauer-Str. 40, 72072 Tübingen, Germany, u.cress@iwm-kmrc.de

Abstract: Research about the effect of cooperative learning settings faces the challenge of dealing with hierarchical data where observations regarding the learners are not stochastically independent. Standard methods like ANOVAs cannot deal with such data adequately. The article introduces multilevel modelling (MLM) as a statistical approach adequate for nested data. MLM allows taking into account interactions between group-level variables and individual-level variables. But MLM requires large sample sizes, and thus many studies fail to adopt MLM. For such studies, it is proposed that some additional statistics be presented.

Introduction

Collaborative learning scenarios are based on the expectation that individuals can take advantage of group processes, and that collaboration and social interaction can facilitate individuals' learning. It is hoped that in a collaborative situation people would learn from each other, would exchange knowledge, and would make use of their array of expertise. The further hope is that bringing people together and having them interact in a certain manner would enable the total group to achieve better results than would be possible for the learners to achieve individually.

With such an optimistic expectation about the efficiency of cooperative learning in mind, we might expect that our statistical repertoire provides adequate methods for analyzing effects of cooperation. But a search for adequate methods comes up empty. We find ourselves at a dead end for adequate methods even for simple research questions. This article provides an alert to pitfalls when using standard statistics for analyzing individuals' behaviour in groups. It explains what multi-level data are, and why they do not provide stochastically independent oberservations. We give a short introduction to the logic of multilevel models. Indeed, multilevel methods need very large sample sizes, which many studies in the context of cooperative learning do not fulfil. So in the conclusion we will make some suggestions how to deal with smaller sample sizes.

Some Pitfalls when analyzing data from collaborative settings

Let us start with a simple example of a prototypical study about collaborative learning: The experimenter would like to compare the efficacy of two different instructions (I1 and I2) for cooperative learning. The experimental approach would require randomly assigning learners to different learning groups, half of them with I1, half of them with I2. Then one would have to measure the learning outcomes. One way of analyzing the data would be to take the individuals as the unit of analysis and pool the students learning with instruction I1 and those learning with instruction I2, without considering that they belong to different learning groups. Then one could compare the two means. But the problem arises that most statistical methods for testing differences require *stochastical independency* of observations. This means that a group member's learning outcome must not depend on the values of the others. But this assumption fully contradicts our expectations, because in cooperative learning we particularly want people to interact and learn from each other.

In collaborative learning, the group members share a *common fate*. When the groups discuss or have any other kind of interaction, then the members of different groups experience different discussions during the collaboration phase. As a consequence, only group members of the same group have equivalent conditions. Not only their common fate, but also the effects of *reciprocal influence* make them stochastically dependent,. Cooperative learning aims to promote active interaction among group members, and so we particularly want them to influence each other reciprocally. This influence is obvious in many collaborative situations. We often observe that a single individual can determine the entire interaction process within a group. Just as a creative group member may stimulate the whole group to have an interesting discussion, an unmotivated member with destructive behaviour can destroy all motivation and any form of discussion among the other group members. In each case, learner behaviour is strongly influenced by fellow group members, and the same individual will behave quite differently according to the group to which he/she belongs. This leads to a hierarchical structure of the data. Figure 1 shows that we have to consider at least two different levels: the group level, describing the different groups, and the individual level, where the individuals are nested within the groups. In such a hierarchical structure of the data we do not expect stochastical independency, because learners of the same groups share a common fate and influence each other reciprocally. And this is what we even want them to do!

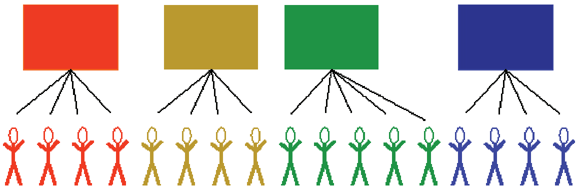

Figure 1. Hierarchical data of cooperative learning settings.

The stochastical non-independence can be measured using *intra-class correlations (ICC)*. This correlation describes the higher (or lower) similarity of individuals within a group compared to the similarity of people belonging to different groups. It is equal to the average correlation between measures of two randomly drawn lower-level units within the same randomly drawn higher level unit. It can also be calculated by the proportion of variance in the outcome variable which is caused by group membership. If the *ICC* in a given data set is significant then it is necessary to deal explicitly with the hierarchical data structure. In this case standard methods such as the OLS-Regression or the standard Analysis of Variance cannot be used. If they are used regardless of a significant *ICC*, then the standard error will systematically be underestimated. An alpha-error inflation thus arises in hierarchical data sets, which leads to significant results which would have not achieved significance in a stochastically independent sample. This means, for example, that due to the low standard error, significance tests do not test against an alpha-error of 5%, as intended by the researcher, but at a much higher alpha-level, depending on the respective *ICC*. Stevens (1996) showed that alpha-error strongly increases with increasing intra-class correlation and group size. For example, in comparing two conditions with a group size of 30 participants and an intra-class correlation of $ICC = .30$, alpha is equal to $\alpha = .59$. This shows that the alpha-error inflation can be enormously high.

So if it is not possible to analyse the data of our example (comparing instructions I1 and I2) on the individual level, one may choose the alternative to take the groups as unit of analysis. For this alternative, one has to aggregate the individual outcomes for each group using the group means for the comparison. The sample size would then be much smaller, because it is now the number of groups and not the number of learners. But this way would also lead us to a dead end. We would be able to compare both instructions on the *group level* (and we would not have any bias here), but this would not allow us to transfer the result to the individual level. Therefore, when the aim of a study is to predict individual learning and not the efficacy of a group as a whole, the problem posed by hierarchical data cannot be solved using aggregated data.

The problem becomes even more complex when we take further individual-level variables into account as predictors or mediators. Let us imagine that we are interested in the effect of prior knowledge on the learning outcome. With a linear regression we would try to predict learning outcome with prior knowledge. We could do this separately for the learners of instructions I1 and I2. But also here we would have different possibilities in calculating the regression:

1. We could base the regression on all learners with I1 respectively I2 without considering that they belong to different learning groups.
2. We could base the regression on the aggregated measures (group means)
3. We could calculate different regressions for each of the groups and compare the groups with I1 and I2.

With stochastical non-independency all three methods can lead to quite different results which cannot be conveyed through simple linear transformation. This is obvious when we have data like those shown in Figure 2 where we compare the first and third possibilities. The left side presents the data of learners with instruction I1, the right side those with instruction I2. The dashed lines presents the regression bases on all learners with instruction I1 respectively with I2 without considering that they belong to different groups. The short lines show the regression base for each group separately. If we wanted to compare the two regression lines based on the individual level data (dashed lines) we would not find any difference between I1 and I2. We would conclude that with both Instructions learners with high prior knowledge learn more. But if we were to focus on the regressions within the groups, we would realize that in the groups with Instruction I2 the learners with higher prior knowledge have lower learning outcomes. So within the groups, I2 leads to a negative correlation between prior knowledge and learning outcome.

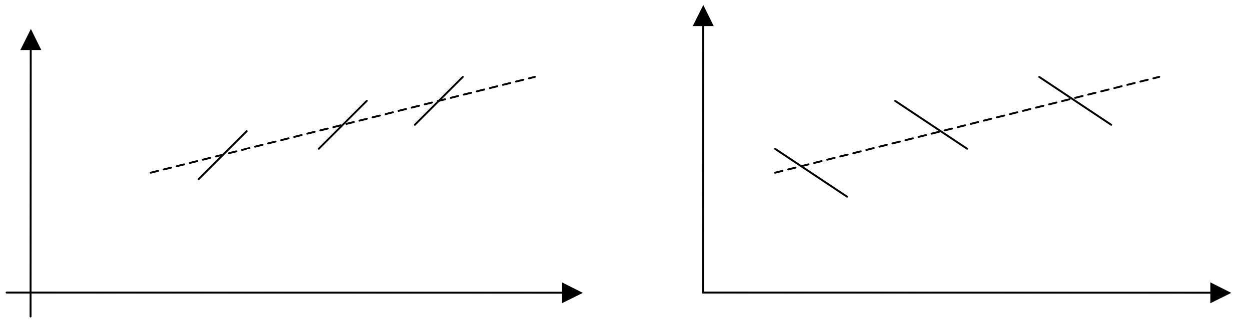

Figure 2. The short regression lines are based on the data within the groups (small regression lines for groups with Instruction I1 on the left side, and with instruction I2 on the right side). The dashed regression lines are based on the data of all learners with Instruction I1 (dashed line left) and with instruction I2 (dashed line right).

Figure 2 illustrates the central problem with data on collaborating individuals: Pooling individual data and handling the data as though they do not come from different groups will lead to results which diverge from data within the groups. These different methods of analysis can lead to quite different regression coefficients as the right Figure 2 shows. Additionally, we have to be aware that the different variations of calculating the regressions rely on different sample sizes. Thus, they would have different degrees of freedom when testing for significance, and regression coefficients of the same size would probably lead to different significance values.

A short introduction to multilevel modelling (MLM)

Figure 2 points the way to dealing with the multi-level problem. A first step is Burstein's slopes-as-outcomes approach (Burstein, Kim & Delandshere, 1989). This method proposes that with hierarchical data a linear regression of a variable *Y* on a variable *X* should pay attention to the slopes resulting from the linear regressions within all the groups. These slopes represent the different covariances of *x* and *y* in the different groups. If these slopes are different, then the group moderates the effect of *X* on *Y*. Taking into account different regression lines for the different groups, this method takes into consideration that the members of one group have equal conditions (and so are stochastically independent) and simultaneously allows different groups to have different conditions. The application of different regression regards the stochastical non-independency of all observations. So Burstein focuses on differences in the slopes and interprets these slopes as an outcome variable for a hierarchical analysis. Different slopes thus show different influences of groups. A visual inspection of Figure 2 would already show that instructions I1 and I2 have different effects on learners' outcomes depending on the instruction and the prior knowledge they have. Indeed, this slopes-as-outcome method does not provide quantitative statistics for the interaction between the individual level predictor prior knowledge and the group level predictor of instruction.

The slopes-as-outcome approach forms the basis of MLM (also called Hierarchical Linear Model), as it was developed by Bryk and Raudenbush in 1992. MLM allows predicting an observation on individual level (*Y*) through predictors on group level (*W*) as well as through predictors on individual level (*X*). With this method the predictor on the group level can be a nominal variable (as in our example before with I1 and I2), but it doesn't have to be a nominal variable. It can also be an interval scale. In order to explain the logic of MLM let us now imagine an example where a learner's outcome (*Y*) is predicted by his/her prior knowledge (*X*) and by his/her group's activity (*W*). Instead of only the one equation of a normal linear regression model, MLM consists of a set of equations which form the linear regression model. The first of its equations (shown in Eq. 1) models the relation between an explanatory variable *X* and a dependent variable *Y* at the lowest level (Level 1).

Eq. 1 $$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$$

Eq. 1 is a standard linear regression, with a regression intercept β_0, a slope β_1 and a residual e_{ij}. But in contrast to normal regression equations, there are two subscripts: the subscript $i = 1,\ldots,n$ refers to the individual and the subscript $j = 1,\ldots,k$ to the different groups. Eq. 1 thus allows differing regression functions with different intercepts and different slopes for each of the *k* groups. This means that β_{0j} and β_{1j} are not constants as in normal regression models, but they are variables and are different for each group *j*.

The variables β_{0j} and β_{1j} are explained by two further equations. These equations describe the processes at level 2. They provide an explanation for the variables β_{0j} and β_{1j} by introducing further explanatory variables at the group level. Such predictors (or explanatory variables) are described by *W*. In our prototype example, we could introduce the groups' activity as such an explanatory variable at the group level. Eq. 2 then describes the linear regression with activity as a predictor of the respective group's intercept, and Eq. 3 describes the linear regression with activity as predictor W of the respective group's slope.

Eq. 2 $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$ Eq. 3 $\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$

These two linear regressions also have intercepts and slopes. These are described using $\gamma_{00}, \gamma_{01}, \gamma_{00}$ and γ_{11}. These gammas are constants with fixed subscripts. Both linear regressions (Eq. 2 and Eq. 3) have residuals $u_{.j}$. They represent the variance which is not explained by the predictor W. The residual is group specific, and in the model u_{0j} and u_{1j} are independent of the residuals e_{ij} at the individual level and have a mean of zero. However, the covariance between u_{0j} and u_{1j} is generally not assumed to be equal to zero.

The full hierarchical linear model thus consists of the three equations Eq. 1, Eq. 2 and Eq. 3. Substituting β_{0j} in Eq. 1 through Eq. 2 and β_{1j} through Eq. 3 results in the following equation:

Eq. 4 $$Y_{ij} = (\gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij}) + (u_{1j} X_{ij} + u_{0j} + e_{ij})$$

Eq. 4 comprises two parts. The first part (first bracket) is fixed (or deterministic), with fixed regression coefficients $\gamma_{00}, \gamma_{01}, \gamma_{00}$ and γ_{00}. The second part (second bracket) is random (also called "error part"). This part reflects the fact that group effects are random and that there is some variance which is not explained by the predictors. With this random part, the model assumes that the groups which are part of the study are a random sample of all possible groups. It is due to this random part that multilevel models are also referred to as "random coefficient models". The term $u_{1j}X_{ij}$ shows that the amount of variance which is not explained by the group predictors can vary across groups. This allows for heteroscendasticity, which is a term for the phenomenon that the variances of the different groups differ. The homogeneity of variances is a necessary pre-condition for the use of many standard methods, and thus heteroscendasticity would not allow for the use of an OLS-regression.

Figure 3 visually presents this hierarchical regression model and visualizes the summands of Eq. 4:

- γ_{00} is the grand mean. It is the learning outcome of an individual in the group with a mean activity ($W = 0$), given that this person has no prior knowledge at all.
- $\gamma_{01}W_j$ represents the influence of activity. The groups with different activity differ in their intercepts. In Figure 3, γ_{01} represents the difference between a person belonging to the group with an activity of $W=1$ and a person of the group with an average activity of W=0, given that these people have no prior knowledge at all.
- $\gamma_{10}X_{ij}$ is the influence of the a student's prior knowledge, the explanatory variable at the first level. It represents the slope of the group with $W = 0$.
- $\gamma_{11}W_jX_{ij}$ represents the cross-level interaction, i.e. the different slopes between the group with an activity of $W=0$ and $W=1$. With a higher W the slope is larger. This means that a group member's prior knowledge has a stronger influence on his/her learning outcome in active groups than in less active groups. Between the homogeneity W and the slope of the linear regression at the first level is a linear relationship.

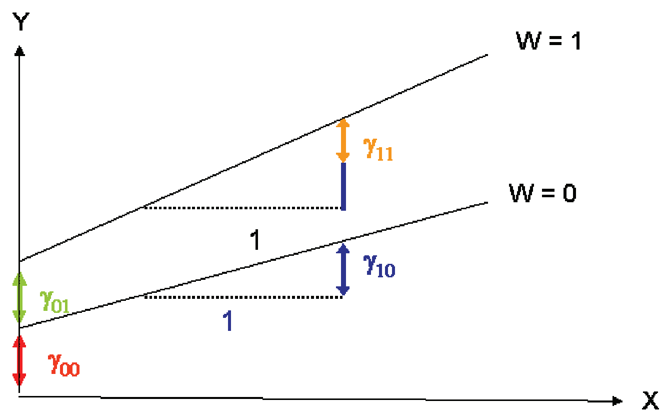

Figure 3. Visual representation of the multi-level model. The regression line with W=0 and W=1 are shown.

For purposes of clarity, the random parts of the model are not visualized in Figure 3, although they will be described verbally.

- $u_{1j}X_{ij}$ is part of the random model and takes into account that the slopes cannot be perfectly predicted for each group, i.e. there is some residual in the prediction. This residual u_{ij} can differ across groups, so that heteroscendasticity (different variances in different groups) is allowed. Standard methods including, for example, ANOVAs do not allow for heteroscendasticity, whereas MLM explicitly deals with and includes it in the model. In Figure 3 this random part of the model would cause that the regression slopes not to be exactly determined by the gammas.
- u_{ij} describes another random part of the model, relating to the residual in the prediction of the groups' regression constants. This means that the explanatory variable at the higher level, W, does not perfectly predict the intercepts and that some unexplained error variance remains. This residual is the same for all individuals of the same group.
- e_{ij} is an individual specific residual showing that not every person's measure directly lies on the individual's respective regression line.

Testing the multilevel model

This full hierarchical model is highly complex. Because of parsimony of theory and data, Hox (2002) suggests that the model be tested using an iterative procedure with five steps.

The *first step* is the intercept-only model (also referred to as "null model" or "empty model"). It includes no explanatory variables at the individual or the group level. The intercept-only model does not explain any variance, but only reveals the proportion of variance caused by the groups. The intercept-only model is given in Eq. 5.

Eq. 5: $Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$

The model is a one-factorial ANOVA with the random factor *u* describing the different groups. This model allows for calculation of the *ICC* which is presented in Eq. 6.

Eq. 6 $$ICC = \frac{Var(u_0)}{Var(u_0) + Var(e_{ij})}$$

In Eq. 6 u_0 describes the between-variance on level 2. Only if the *ICC* is significant must a multilevel model be used.

The *second step* includes the lower-level explanatory variable *X* as fixed variable (i.e. the variance components of the slopes are constrained to zero). This results in the following ANCOVA model with the covariate *X* and a random group factor *u*:

Eq. 7: $$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{0j} + e_{ij}$$

If this model has a significantly better fit than the intercept-only model (which can be tested using a chi-square test), then in a *third step* a model can be chosen which includes the explanatory variables at the group level.

Eq. 8: $$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_{1j} + u_{0j} + e_{ij}$$

The *fourth step* allows for varying slopes in the different groups, as so it is also called "random coefficient model".

Eq. 9: $$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_{1j} + u_{1j}X_{ij} + u_{0j} + e_{ij}$$

In the *fifth step* also a cross-level interaction between the explanatory group level variable W and the individual level explanatory variable X is introduced. This enables the different slopes of the groups to be predicted by the group level explanatory variable.

Eq. 10: $$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_{1j} + \gamma_{11}W_{1j}X_{ij} + u_{1j}X_{ij} + u_{0j} + e_{ij}$$

This iterative procedure demonstrates that even when data result from a hierarchical structure, it may not always be necessary to use the full hierarchical model shown in Eq. 4. Less complex models with fewer coefficients are often sufficient. But if we find a significant ICC then we have to determine if one of those models is necessary.

The model described thus far is a complex model with two levels and one explanatory variable for each level. According to the experimental design, larger or smaller models can also occur. For example, the appropriate equation for a 2-level model which does not include any explanatory variables at the lower level would be:

Eq. 11:
$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + e_{ij}$$

This model is an ANOVA model with a random effect and can also be calculated using standard software such as SPSS. In our example such a model would be appropriate if we would like to provide a model for the different groups' different effects on the learning outcomes and if we would like to predict these effects with the group's activity W.

Hierarchical models in research about collaborative learning

Over the course of the last several years, multilevel models have become more common. A search for the terms "multilevel" or "HLM" in the database PsychInfo reveals that the very first articles appeared in the eighties and that the number of articles has greatly increased to more than 350 over the last five years. In modern educational psychology, hierarchical methods have especially gained a strong position through large-scale studies in the context of evaluating educational systems. In studies such as OECD-PISA which compare educational systems in different countries, it is obvious that data are nested (learners in classes, classes in schools, and schools in school systems or in countries). In the area of collaborative learning MLM was used very seldom. This is caused by the fact that studies about cooperative learning normally do not have large enough samples that would be necessary for applying MLM. In her simulation studies, Kreft (1996) states that a two-level model requires approximately 30 groups of 30 individuals, 60 groups of 25 individuals or 150 groups of 5 individuals in order to test for cross-level interaction with adequate power. An adequate study should therefore be based on a minimum of approximately 1,000 individuals. Kraft found a rapid decrease in statistical power when the sample size fell below this threshold, and found a high risk of failure to detect existing cross-level interaction effects. In their simulation studies, Maas and Hox (2005) found evidence that such enormous sample sizes are not needed. But they state that especially a small sample size in level two (fewer than 50 groups) leads to biased estimates of second-level standard estimates. In simulations with only 10 groups they found a bias up to 25%.

This represents a problem within research in collaborative learning, where sample sizes are for the most part considerably smaller. With small sample sizes at the group level, the potential for detecting group level effects and the confidence of the estimated regression coefficient values are low. Nevertheless, some authors have begun to use multilevel models for analyzing cooperative learning despite small sample sizes. Strijbos, Martens, Jochems, and Broers (2004) investigated the effect of roles on group effectiveness in CSCL with 10 groups of approximately 4 learners each. Strijbos, Martens, Jochems and Boers (2007) used a sample of 13 groups. Piontkowski, Keil and Hartmann (2006) studied the effect of a sequencing chat tool based on the participation of 40 groups of three learners each. All three studies found significant intra-class correlations (ICC between .32 and .45) and were able to explain some of the group variance using second level factors. More complex models with repeated measurements were used by Schellens, Van Keer and Valcke (2005) for predicting learners' knowledge construction in asynchronous discussion groups. Data were collected on four measurement occasions (according to four discussion themes) for each of the 286 students, who were nested in 23 groups. The 3-level hierarchical model revealed a significant influence of the student-level predictors (attitude toward the learning environment and engagement in the discussion groups), but no group-level effects. The follow-up study of De Wever, Van Keer, Schellens and Valcke (2007) has a similar 3-level design. Their data sets consist of fourteen 10-person groups, with 4 measurement occasions each. This study confirmed the results of the previous one in revealing no significant group effect. Schellens, Van Keer, Valcke & De Wever (2007) assigned 230 students to 23 asynchronous learning groups to test the influence of student, group and task characteristics on students' final exam scores and their levels of knowledge construction. It revealed that only 6% of the overall variability in the final exam scores is explained by the group characteristics. With regard to knowledge construction the situation was different. Here about 19% of the variance was explained by differences among groups. Students in groups which were active in discussion performed at a qualitatively higher level than those belonging to less active groups. Chiu and Khoo (2003; 2005) analyzed the effect of rudeness and status on group-problem-solving with 80 people belonging to 20 groups. They found significant effects of the group level which explained 12% of the total variance.

In sum, it seems too early to summarize the results of these studies. But it appears that the amount of variance explained by groups is sometimes rather small. In many of these studies the use of MLM could be criticized as inadequate in the case of such small samples sizes on the highest level. But as long as there are no better methods for small groups, MLM seems an important way for estimating the group influences and interaction between group and individual predictors. Research in collaborative learning implicitly assumes that collaboration of learners has an effect, but the data do not always support this assumption, MLM would be a potent method for testing this assumption. But so far we do not have a clear picture about the biases MLM produces with small samples. For future research it would seem desirable to apply different statistical means in order to be able to compare their results. In its current state, research is only at the very beginning of a discussion of methodological issues for measuring the effect of collaboration and of establishing an adequate methodology (Snijders & Fischer, 2007). Given that no satisfying solution for the multilevel problem in small groups has thus far been found, studies with much smaller samples sizes and critical discussion of those studies may help to widen the focus of research in collaborative learning and further direct attention to concurrent existing deficits in its methodology.

Conclusion and suggestions

Since CSCL research is explicitly founded on the claim that learning in groups can improve individual learning processes and enhance learning outcomes, it is essential that efforts be made to find a method which is adequate for testing and identifying such effects. Recent research has often been restricted to traditional methods which are not able to deal with the specific requirements of the complex data resulting from cooperative learning. Some authors are aware of the multilevel problem and subsequently have decided to analyze the processes solely at the group level using exclusively aggregated data. This method is too superficial, however, when it comes to analysing the complex combination of individual processes and group influences involved in CSCL settings. Using groups as the unit of analysis is a waste of data and reduces quantitative analyses to a comparison of different instructions or learning settings without considering that learning is an individual process which, while taking place in a group, is primarily an individual cognitive process. It is precisely the analysis of this interaction between group influences and individual prerequisites which should constitute an important goal within CSCL research.

While a consideration of groups as units of analysis is unsatisfying, it is not acceptable to neglect the hierarchical structure of the data and analyze the individual data at the individual level without considering group effects. As shown in our examples, this yields misleading results. Data can only be analyzed at the individual level given that no significant intra-class correlation exists. This in turn, however, also means that the group has no effect. In dealing with hierarchical data, the use of MLM is adequate. Intra-class correlations can be used to identify the effect of collaboration, and factors of the learning environment (instruction, tools, roles, content etc.) can be interpreted as mediators and included in a hierarchical linear model as second level predictors. Even if MLM appears to be the optimal method for analyzing collaborative situations, the fact that it requires large sample sizes can be a hindrance.

As long as no optimal statistical methods exist for the analysis of small sample sizes, research about collaborative learning should continue attempting to use multilevel models, even if they are imperfect. As a minimum standard, the *intraclass correlation* (*ICC)* should be calculated and tested for significance, whenever the sample size is large enough. If a learning setting does not produce a significant intra-class correlation, then the groups do not appear to have a systematic impact on people's learning. There may be an influence on the learners in single groups, but in this case influence remains unpredictable by variables describing the group.

In the case of a significant *ICC,* the slopes of the different groups can be compared, if the study includes an individual-level predictor. If a study includes one or more group-level predictors, then the data can be analyzed with a random-coefficient model (ANOVA with varying instead of fixed factors), given that the groups' different intercepts are of interest. All of these methods can be used with smaller sample sizes and are adequate for many CSCL studies which do not apply a full hierarchical design with individual level predictors, group level predictors and cross-level interactions.

In general, CSCL research should address the hierarchical structure of its data in a more explicit manner. We might change our point of view so as not to interpret groups only as a source of unintended error variance, but we should also be interested in group effects and cross-level interactions as important outcome variables.

References

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Burstein, L., Kim, S.S., & Delandshere, G. (1989). Multilevel investigations of systematically varying slopes: Issues, alternatives, and consequences. In D. Bock, (Ed.), *Multilevel Analysis of Educational* Data (pp. 233-279). San Diego: Academic Press.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions. *Journal of Educational Psychology, 95*, 506-523.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multilevel modeling. *Small Group Research, 36*, 600-631.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Leaning & Instruction, 17*, 436-447.

Hox J.J. (2002). *Multilevel analysis: techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kreft, I. (1996). *Are multilevel techniques necessary? An overview, including simulation studies.* Retrieved March, 20, 2007 from http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Maas, C. J. M., & Hox, J. (2005). Sufficient samples sizes for multilevel modeling. *Methodology, 1,* 86-92.

Piontkowski, U., Keil, W., & Hartmann, J. (2006). Analyseebenen und Dateninterdependenz in der Kleingruppenforschung am Beispiel netzbasierter Wissensintegration. *Zeitschrift für Sozialpsychologie, 37,* 41-50.

Schellens, T., Van Keer, H., & Valcke, M. (2005). The impact of role assignment on knowledge construction in asynchronous discussion groups: A multilevel analysis. *Small Group Research, 36,* 704-745.

Schellens, T., Van Keer, H., Valcke, M., & De Wever, B. (2007). Learning in asynchronous discussion groups: A multilevel approach to study the influence of student, group and task characteristics. *Behaviour & Information Technology, 26*, 55-71.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed). Lawrence Erlbaum, Mahwah, NJ.

Strijbos, J. W., & Fischer, F. (2007). Methodological challenges in collaborative learning research. *Learning & Instruction, 17,* 389-393.

Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of funcitonal roles on group efficiency: Using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 25,* 195-229.

Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: a matter of triangulation. *Computers in Human Behavior, 23,* 353-380.