Running head: MULTILEVEL ANALYSIS IN CSCL RESEARCH

MULTILEVEL ANALYSIS IN CSCL RESEARCH

Jeroen Janssen[*] *Gijsbert Erkens*[*], *Paul A. Kirschner*[**] , *Gellof Kanselaar*[*],

[*] Research Centre Learning in Interaction, Utrecht University, The Netherlands

[**] Centre for Learning Sciences and Technologies, Open University, The Netherlands

INTRODUCTION

CSCL researchers are often interested in the processes that unfold between learners in online learning environments and the outcomes that stem from these interactions. However, studying collaborative learning processes is not an easy task. Researchers have to make quite a few methodological decisions such as how to study the collaborative process itself (e.g., develop a coding scheme or a questionnaire), on the appropriate unit of analysis (e.g., the individual or the group), and which statistical technique to use (e.g., descriptive statistics, analysis of variance, correlation analysis). Recently, several researchers have turned to multilevel analysis (MLA) to answer their research questions (e.g., Cress, 2008; De Wever, Van Keer, Schellens, & Valcke, 2007; Dewiyanti, Brand-Gruwel, Jochems, & Broers, 2007; Schellens, Van Keer, & Valcke, 2005; Strijbos, Martens, Jochems, & Broers, 2004; Stylianou-Georgiou, Papanastasiou, & Puntambekar, chapter #). However, CSCL studies that apply MLA analysis still remain relatively scarce. Instead, many CSCL researchers continue to use 'traditional' statistical techniques (e.g., analysis of variance, regression analysis), although these techniques may not be appropriate for what is being studied. An important aim of this chapter is therefore to explain why MLA is often necessary to correctly answer the questions CSCL researchers address. Furthermore, we wish to highlight the consequences of failing to use MLA when this is called for, using data from our own studies.

MULTILEVEL ANALYSIS: A 'NEW' METHODOLOGICAL APPROACH IN CSCL RESEARCH

Over the last 5 years or so, multilevel analysis (MLA) has been adopted by several CSCL researchers to answer their research questions, because MLA is especially suited to "appropriately grasp and disentangle the effects and dependencies on the individual level, the group level, and sometimes the classroom level" (Strijbos & Fischer, 2007, p. 391). Although MLA is a relatively 'new' technique, especially to CSCL researchers, its development already began in the 1980's (Snijders & Bosker, 1999) and is since then used in several research disciplines.

MLA was initially embraced by educational researchers interested in school effectiveness research because it is well suited to the type of datasets they analyze. Consider, for example, an educational researcher interested in the effect of class size (i.e., the independent variable) on student achievement (i.e., the dependent variable). To investigate this effect, he or she would collect data about class sizes in different schools, as well as data on student achievement (e.g., standardized test scores on language, mathematics, and so on). However, such a research question poses several problems. First, this research question yields a *hierarchically nested dataset* with students nested within classrooms, and with classrooms nested within schools (and if this was an international study, even with schools nested within countries). Furthermore, the researcher would encounter the problem of *nonindependence* of his/her dependent variable of interest. Many statistical techniques (e.g., *t*-test, analysis of variance, regression analysis), assume that the achievement scores (or other dependent variables) of the students in the dataset are independent from each other (Hox, 2003; Kashy & Kenny, 2000; Snijders & Bosker, 1999). In the example just provided, this will probably not be the case. Due to a common experience for example (e.g., the teaching they receive by their teacher), the scores within one classroom may not be independent at all since the overall classroom environment will affect all children in the class and even the behavior of individuals in the class will affect the others. Finally, the imaginary educational researcher

would have to take into account that his or her variables of interest, class size and student achievement, are measured at different levels. Class size is measured at the class level, while achievement is measured at the student level. The number of available observations for both variables differs (i.e., the number is smaller for class size than for achievement). To properly address these issues, MLA was developed, and since then it became an important technique for school effectiveness research (Bosker & Snijders, 1990; De Leeuw & Kreft, 1986).

Social psychologists have also acknowledged the analytical problems described above. They are frequently interested in how individuals' thoughts and behaviors are influenced by other people. Many social psychological concepts involve two or more persons (e.g., attraction, interactive behavior, marital satisfaction) and thus the behavior of individuals *within a group* is often the focus of study (Kashy & Kenny, 2000). A social psychologist might for example be interested in how the division of household chores (i.e., the independent variable) affects marital satisfaction (i.e., the dependent variable). To answer this question, the researcher would have to observe married couples and record who did which chore (and calculate for example a ratio), and to administer a questionnaire to both spouses to measure their marital satisfaction. From this example it becomes clear the social psychologist encounters the same problems as the educational researcher does. Both encounter the problem of hierarchically nested datasets (in this case individuals nested within couples), both involve variables at different levels of measurement (in this case the household chores ratio is measured at the level of the couple, whereas marital satisfaction is measured at the level of the individual), and in both cases the observations of the dependent variable are probably not independent (in this case there might even be a negative relationship: the husband may be more satisfied if he does little housework, while this may negatively affect the marital satisfaction of his wife).

The problem of statistical nonindependence of dependent variables (i.e., group members exerting a psychological influence on each other) has received considerable attention in social psychology since the 1980's (e.g., Bonito, 2002; Kenny, 1995, 1996; Kenny & Judd, 1986; Kenny & Judd, 1996) but less so in CSCL research (notable exceptions are for example, Cress, 2008; Strijbos et al., 2004; Stylianou-Georgiou et al., see chapter #). It is therefore not surprising that social psychologists frequently use MLA to deal with these issues (cf., Bonito & Lambert, 2005; Campbell & Kashy, 2002; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002), while this technique is less often used in CSCL research.

THE PROBLEMS CSCL RESEARCHERS ENCOUNTER

Similar to other research disciplines, CSCL researchers encounter the abovementioned problems of hierarchically nested datasets, nonindependence of dependent variables, and differing units of analysis. We explain these problems below.

Hierarchically Nested Datasets

In CSCL-environments, students work in groups. Studying online collaboration therefore often involves investigating group processes and how these processes are affected by contextual factors (e.g., the environment itself, the composition of the group, prior knowledge and experiences of the group members). It is not difficult to understand this leads to *hierarchically nested datasets*, since groups consist of two or more individuals and thus in these cases individuals are nested within groups. In many cases, CSCL researchers will encounter at least two levels: the group and the individual. The group is then the macro- or level-2 unit and the individual the micro- or level-1 unit (Hox, 2003; Snijders & Bosker,

1999). CSCL researchers may also use datasets that have even more levels of analysis. A researcher might for example be interested in the effects of the teacher's experience with CSCL on the way his or her students collaborate online. This researcher will have a dataset with three levels: students are nested within groups, while groups are nested within teachers' classrooms. Another CSCL researcher might be interested in the development of students' online interactive behavior over time. This researcher would therefore collect data about students' interactive behavior on different measurement occasions. This would also lead to a dataset with three levels: measurement occasions are nested within students, and students are nested within groups (Kenny, Kashy, & Cook, 2006; Snijders & Bosker, 1999). Whenever researchers encounter datasets with hierarchically nested data, MLA is needed to appropriately model this data structure since it can appropriately disentangle the effects of the different levels on the dependent variable(s) of interest (Snijders & Bosker).

Nonindependence of Dependent Variables

Because their participants work in groups, CSCL researchers also encounter the problem of *nonindependence* of their dependent variables (Cress, 2008). This means students within a group may be more similar to each other than are persons from different groups (Kenny et al., 2002). In the case of CSCL, the main source of this nonindependence is the mutual influence group members have on each other (Bonito, 2002; Kenny, 1996). In our own studies, which we describe in detail later on in this chapter, students could discuss with each other through a Chat-window and a Forum. Through these discussions, students influenced each other. In some cases for example, a student displayed negative behavior, and this prompted the other group members to respond negatively as well. Furthermore, some students in our studies were very active in the Chat conversations (e.g., they proposed a lot of strategies and asked a lot of questions). This could have triggered the other group members to

also become more active in the chat as well. Such an influence of students on their group members' communication and behavior is nearly always present in CSCL research, because in CSCL-environments students communicate and collaborate to solve complex problems (Kreijns, Kirschner, & Jochems, 2003).

This reciprocal influence of group members is not necessarily positive, it can also be negative. In the previously mentioned example concerning active group members stimulating other group members to be more active, the reverse could also happen: When one group member is very active in the learning environment, this may trigger other group members to "sit back" and do little since that other group member is doing so much (O'Donnell & O'Kelly, 1994; Webb & Palincsar, 1996). Kenny et al. (2002) therefore noted that mutual influence can not only cause students to behave more similarly, but may also cause students to behave differently from their group members. This is called the *boomerang effect* (Kenny et al., 2006). Another example is that when group members behave negatively, a student may decide to counter this by displaying more positive behavior. Role assignment (cf., Schellens et al., 2005; Strijbos et al., 2004; Strijbos, Martens, Jochems, & Broers, 2007) may also lead to differential behavior. If one group member, for example, is given the task to ask critical questions, while the other group member has to monitor task progress, this may lead to differing behavior (e.g., the first student will ask many questions, but will display less metacognitive behavior, while the second student may display high levels of metacognitive behavior but may ask fewer questions). Kenny et al. therefore make a distinction between *positive nonindependence* where group members influence each other in such a way that they behave more similarly and *negative nonindependence* where group members influence each other to behave differently. Thus, since group members influence each other in a group context, this will likely lead to either positive or negative nonindependence of the dependent variables that are being investigated which in turn has to be dealt with during data analysis.

The degree of nonindependence can be estimated using the *intraclass correlation coefficient*[1] (ICC, cf., Kashy & Kenny, 2000; Kenny et al., 2002). Values of the ICC can range from -1 to +1. An ICC of +1 for satisfaction with the collaborative process (scored on a 4-point scale ranging from 1 to 4) for example, indicates that when a group member has a score of 4 on this measure, the other group members will also have a score of 4. Conversely, an ICC of -1 for the same measure indicates that when one student has a score of 4 on this measure, his or her partners will have a score of 1.

An alternative interpretation of the ICC is in terms of the *amount of variance that is accounted for by the group* (Kenny et al., 2006). When the ICC for satisfaction with the collaborative process is found to be .40 for example, this means the 40% of the variance in this measure is accounted for by the group, and thus that 60% is accounted for by other (e.g., individual) factors.

The dependent measures that CSCL researchers are interested in will often be non-independent (Cress, 2008). Strijbos et al. (2004) for example, studied the effect of roles on perceived group efficiency. They found an ICC of .47, meaning 47% of this measure is accounted for by the group. Group members displayed rather similar levels of perceived group efficiency, probably due to their common experiences in the CSCL environment. In a related study, Strijbos et al. (2007) found a similar influence of group level factors on group members' individual perceptions (ICC = .45). In the two studies described in the chapter by Stylianou-Georgiu et al. (chapter #), group level factors explained 11% and 8% of the variance respectively. On the other hand, not all researchers find similar substantial amounts of variance accounted for by the group. De Wever et al. (2007) for example, report only 3% of the students' level of knowledge construction was linked to the group level. However,

[1] For an excellent description on how to compute the ICC for a specific dataset, the reader is referred to Kenny et al. (2006).

these examples still illustrate the presence of nonindependence in datasets of CSCL researchers.

Nonindependence needs to be addressed when conducting statistical analyses, because it distorts estimates of error variances, thus making standard errors, *p*-values, and confidence intervals invalid when this distortion is not taken into account (Kenny, 1995; Kenny et al., 2006). Traditional statistical techniques such as *t*-tests, analyses of variance, and regression analyses cannot cope with this distortion because they assume the variables are independent. Therefore CSCL researchers using these types of analyses run an increased risk of committing Type I or Type II errors (Kashy & Kenny, 2000). Whether the chance to falsely reject (Type I error) or falsely accept (Type II error) the null hypothesis is increased, depends on the sign of the ICC (either positive or negative), and the type of dependent variable for which the ICC was calculated (see Kashy & Kenny for a detailed discussion).

Like any correlation coefficient, the ICC can be tested for significance. When the ICC is significant, its effect is large enough to bias statistical tests as described above (Kenny et al., 2006). However, because sample sizes are often small in CSCL research, the ICC may not be significant, while it is actually large enough to bias standard errors, *p*-values and so on. Kenny et al. (2002) therefore propose assuming group data are nonindependent even though the ICC is not significant.

Differing Units of Analysis

A final problem that CSCL researchers encounter concerns the *differing units of analysis* their datasets often contain. This has to do with the abovementioned hierarchical structure of their datasets. Some variables that CSCL researchers are interested in are measured at the individual level (e.g., gender, interactive behavior, familiarity with other group members), whereas other variables are measured at the group level (e.g., gender group

composition, group performance, group consensus). Savicki and Kelly (2000) for example, studied the effect of gender and gender group composition (male-only, female-only, or mixed) on satisfaction with online collaboration. Their dependent variable was measured at the individual level (satisfaction), while their two independent variables were measured at the both the individual (gender) and group (gender group composition) level. Thus, their dataset contained variables with differing units of analysis.

Another example comes from a study conducted by Schellens, Van Keer, Valcke, and De Wever (2007). During their study, students collaborated in asynchronous discussion groups. They were interested in the impact of individual variables (e.g., gender, learning style) and group variables (active group versus relatively inactive group) on students' final exam scores. Their analyses therefore included independent variables measured at both the individual and the group level, while their dependent variable was measured at the level of the individual student.

To be able to cope with the different units of analysis encountered by Savicki and Kelly (2000) and Schellens et al. (2007), MLA is needed, because traditional statistical techniques cannot properly take these differing units of analysis properly into account (Hox, 2003; Snijders & Bosker, 1999).

COMMON ANALYSIS STRATEGIES

In this section we describe three strategies that researchers can use to deal with the data analytical problems described in the previous sections, namely ignoring nonindependence of dependent variables, aggregating or disaggregating data, and MLA.

Ignoring Nonindependence

A first strategy, and also still the most common practice during the analysis of group data (Kenny et al., 2002), is to ignore the hierarchical structure of the dataset, the nonindependence, and the differing units of analysis and perform statistical techniques such as *t*-tests or (M)ANOVA's (Cress, 2008). As we discussed previously, this biases significance tests of inferential statistics (e.g., *t* or *F*-values), sometimes making tests too liberal and, and at other times, too conservative.

Ignoring nonindependence is a frequently encountered approach in CSCL research. Francescato et al. (2006) for example, studied differences in students' evaluations of collaboration in online and face-to-face learning groups. Among other things, they investigated whether online learning groups perceived differing levels of social presence and satisfaction with the collaborative process than face-to-face groups. However, they found no differences between online and face-to-face groups using analyses of variance, although there was a tendency for online groups to be more satisfied with the collaborative process ($p = .17$). Because this study involves students working in groups, the evaluations of Francescato et al.'s students are most likely nonindependent. However, their analyses fail to take nonindependence into account, and thus the *p*-values reported by the authors might be biased. This could lead to a false acceptance of the null hypothesis (i.e., no differences between face-to-face and online learning groups). Using a more appropriate statistical technique, MLA, Francescato et al. might have been able to demonstrate significant differences between face-to-face and online learning groups.

Another example comes from the work of Guiller and Durndell (2007) who studied the effect of gender on students' linguistic behavior in online discussion groups. Guiller and Durndell studied whether male more absolute adverbials (i.e., strong assertions such as 'obviously') and imperatives (i.e., giving commands) than female students. In order to answer

this question they coded students' messages and classified each message in terms of the linguistic behavior shown by the students. Guiller and Durndell then used χ^2-analyses to determine whether male and female students differed with respect to these behaviors. Although the authors found male students to use more absolute adverbials and imperatives, the corresponding χ^2-values were not significant. However, by using χ^2-analyses they too ignored the nonindependence of their dependent variables. Again, group members communicated and discussed with each other, so therefore they likely influenced each other. Using MLA, Guiller and Durndell, might have been able to detect statistically significant differences between male and female students on use of certain linguistic behaviors.

Aggregating or Disaggregating Data

Another strategy to deal with the problems described in the previous section is to *aggregate* individual data to the level of the group (Snijders & Bosker, 1999). This involves summing the scores of the individual group members to create an aggregated group score.

This strategy is used in a study described by Van der Meijden and Veenman (2005). Van der Meijden and Veenman compared dyads using face-to-face (FTF, $N = 20$) and computer-mediated communication (CMC, $N = 22$) with respect to exchange of high-level elaboration (e.g., elaborate explanations or requests for help). Students' collaboration was coded using a coding scheme. However, the percentages of high-level elaboration "were calculated by summing the individual code frequencies" (p. 843) and dividing these by the total number of utterances. An independent samples *t*-test was then used to establish whether FTF and CMC conditions differed significantly with respect to high-level elaboration. High-level elaboration was thus treated as a group level variable. Such an analysis however, ignores the fact that high-level elaboration is in essence an individual level variable (although it may be affected by group level variables). Furthermore, by aggregating to the group level,

this analysis uses fewer observations for high-level elaboration than are available. For this variable only 20 + 22 = 42 observations are used, while in effect there are 42 * 2 = 84 observations. Therefore Van der Meijden and Veenman run the risk of committing a Type II error. Fortunately, in their study the differences between FTF and CMC were large enough to detect a significant difference between FTF and CMC groups with respect to the percentage of high-level elaborations exchanged.

The reverse strategy can also be applied: treating group level data as if they were measured at the individual level. This is called *disaggregation*. Consider for example, the study by Savicki, Kelley, and Lingenfelter (1996) about the effects of gender group composition on students' satisfaction with the collaborative process. Group composition was measured at the group level (all male, all female, or mixed groups), while satisfaction was measured at the individual level (students completed a questionnaire individually). In total, their sample consisted of 6 groups and 36 students. Savicki et al. conducted an analysis of variance to examine whether group composition affected satisfaction. However, this analysis does not take into account that group composition was measured at the group level. Thus, Savicki et al.'s analysis uses 36 observations for the group composition variable, while in fact there are only 6 observations for this variable. This led to an exaggeration of the actual sample size for this variable and increased the chance of committing a Type I error (Snijders & Bosker, 1999).

Multilevel Analysis

MLA was designed specifically to cope with hierarchically nested data (Hox, 2003; Snijders & Bosker, 1999). Furthermore, it is a useful technique when researchers use datasets that have different units of analysis, such as group and individual level variables (Kenny et

al., 2006). Finally, MLA can deal with the nonindependence of observations that results from the mutual influence group members have on each other (Snijders & Bosker).

At present, MLA is slowly finding its way to the CSCL research community: more and more CSCL researchers are using MLA to analyze their data. In the previously mentioned study of Strijbos et al. (2004), two conditions were present: a condition in which specific roles (e.g., project planner, editor) were assigned to students and a condition without role assignment. Thus, condition was a group level independent variable. Perceived group efficiency was measured using several questionnaires, and was therefore an individual level dependent variable. It is not difficult to see that in the Strijbos et al. study hierarchically nested data were collected since students were nested in groups. Furthermore, their study employed variables measured at different units of analysis. Finally, as we previously mentioned, nonindependence was present in their dataset, since they reported an ICC of .47 for perceived group efficiency. Strijbos et al. therefore constructed a ML model with perceived group efficiency as dependent variable and condition (role or non-role assignment) as an independent variable. Using MLA, they were able to model the nonindependence in their datasets and to analyze their dependent and independent variable at their appropriate levels of analysis.

ILLUSTRATION OF PROBLEMS AND ANALYSIS STRATEGIES

In this section we will illustrate more elaborately the three problems (hierarchically nested datasets, nonindependence, and differing units of analysis) and strategies for data analysis (ignoring nonindependence, aggregating or disaggregating, and MLA) that were

described in the previous sections. In order to do so, we utilize data from two different studies we conducted to illustrate three examples.

Example 1: Impact of an awareness tool on online discussion

The first example comes from the data collected for a study described in Janssen, Erkens, and Kanselaar (2007). For this study we developed an awareness tool (cf., Engelmann, Dehler, Bodemer, & Buder, 2009), called the Shared Space, which visualized the amount of agreement or discussion among group members during online synchronous chat discussions. We hypothesized that giving students such an awareness tool, would raise their awareness about the way they conducted their online discussions. In one condition students used the Shared Space (SS) to communicate online, while in the other condition (No SS) the students communicated through a regular chat-tool. We examined - amongst others - the effect of experimental condition on the number of times students evaluated the social interaction positively during their online conversations.

During this study we encountered the three abovementioned problems. First, because in this study students worked in groups, we had a *hierarchically nested dataset*. Furthermore, we found an ICC of .41 for our dependent variable, indicating a considerable influence of the group on this variable and the presence of *nonindependence*. In this study we also encountered the problem of *differing units of analysis*. Our dependent variable, the number of positive evaluations communicated by the students, was measured at the level of the individual. Because the group as a whole was assigned to either the SS or No SS condition, our independent variable, experimental condition, was measured at the level of the group. Moreover, we also wanted to control for students' level of participation, because some students were more active in the online discussions than others. Thus, our analysis also included a covariate, also measured at the level of the individual.

As we described before, we have three options when analyzing our data. If we chose the first option, *ignoring nonindependence*, we could use regression analysis to answer the question whether the Shared Space had an effect on the number of times students evaluated the collaboration positively. In this regression analysis, we include number of positive evaluations of the collaboration a student typed in the Chat-tool as a dependent variable and condition (effect coded with Shared Space as +1 and No Shared Space as -1) as an independent variable. Furthermore, we also include participation (e.g., the total number of messages students sent) in the regression equation to control for the fact that some students were more active during the online collaboration than others. As can be seen in Table 1, we find no effect of condition (Shared Space or No Shared Space) using this regression model on positive evaluations of the collaboration, $B = 0.20$, $SE = 0.14$, $p = .08$ (one-tailed significance). Thus, if we adopt a strategy that ignores nonindependence, we would conclude that the Shared Space does not influence the number of positive evaluations of their collaboration typed by students.

Table 1: Regression analysis of the effect of the Shared Space on number of positive evaluations of the collaborative process typed in the Chat-tool.

	B	*SE B*	*β*
Condition (-1 = No SS, +1 = SS)	0.202	0.142	.131
Participation	0.001	0.001	.140

If we chose the second option, namely to *aggregate* our data, we could calculate the sum of positive evaluations of the collaboration for each group. On average, Shared Space groups exchanged 2.25 ($SD = 3.16$) positive evaluations of the collaboration, while No Shared Space groups exchanged only 1.00 ($SD = 1.95$) of these messages. Again, we could then use a regression analysis to examine the effects of condition (Shared Space or No Shared Space) on the number of positive evaluations of the collaborative process exchanged by the

group. This regression analysis includes number of positive evaluations as the dependent variable, condition as the independent variable, and again level of participation (i.e., the total number of messages sent by the whole group) as a control measure. The results of the regression analysis are displayed in Table 2. As can be seen, condition was not found to have a significant impact on the number of positive evaluations of the collaborative process sent, $B = 1.20$, $SE = 0.83$, $p = .06$. This yields a conclusion comparable to the previously described strategy of ignoring nonindependence: the Shared Space does not have an influence on the amount of positive evaluations of the collaborative process exchanged during online collaboration.

Table 2: Regression analysis of the effect of the Shared Space on number of positive evaluations of the collaborative process typed in the Chat-tool by the group.

	B	*SE B*	*β*
Condition (-1 = No SS, +1 = SS)	1.202	0.830	.228
Participation	0.001	0.001	.166

Our final option is to use *multilevel analysis* to study the effects of the Shared Space on students' use of positive evaluations of the collaboration. In our study, we constructed a ML model that included number of times a student typed a positive evaluation of the collaboration as a dependent variable and condition (Shared Space or No Shared Space) as an independent variable. Furthermore, we included participation (e.g., total number of messages sent) again to control for the fact that some students typed more messages than other students. As can be seen in Table 3, we found a significant effect of the Shared Space on the number of positive evaluations of the collaboration students typed, $\beta = 0.20$, SE = .14, $p = .04$ (one-tailed significance). Although the differences in *p*-values are small (see Table 4) and the differences may not seem spectacular, this last analysis strategy leads to a different conclusion than ignoring nonindependence or aggregating data, namely that the Shared Space

affects the number of positive evaluations of the collaboration. Thus, in this case MLA prevented us from making a Type II error (i.e., falsely accepting the null hypothesis).

Table 3: Multilevel analysis of the effect of condition (Shared Space or No Shared Space) on number of positive evaluations of the collaborative process exchanged.

	β	*SE*
Participation	0.01*	0.00
Condition (SS or No SS)	0.20*	0.14
Deviance	429.14	
Decrease in deviance	2.06*	

* $p < .05$.

Table 4: Summary of differing results for effects of Shared Space on students' positive evaluations of the collaborative process.

	Ignoring nonindependence	Aggregating data	Multilevel analysis
Statistical analysis	Regression analysis	Regression analysis	MLA
Significance of effect of condition	Not significant, $p = .08$	Not significant, $p = .08$	Significant, $p = .04$
Conclusion	No effect of Shared Space on positive evaluations of collaborative process	No effect of Shared Space on positive evaluations of collaborative process	Positive effect of Shared Space on positive evaluations of collaborative process

Example 2: Influence of representational guidance on student learning

Our second example comes from a study reported in Janssen, Erkens, Kirschner, and Kanselaar (in press). In this study we investigated the effects of representational guidance (cf., Suthers, 2001; Suthers & Hundhausen, 2003) on students' performance on a knowledge post-test. Our design used two conditions: In one condition students used a Graphical Debate-tool to construct external representations of a historical debate, while in the other condition students used a Textual Debate-tool to construct such representations. Both versions of the tool differed with respect to the representational guidance they offered to the students. The Graphical Debate-tool made extensive use of visualization techniques to visualize certain

aspects of the collaborative problem solving process (e.g., Was there a balance between the number of arguments pertaining to both positions?). We hypothesized that the representational guidance offered by the Graphical Debate-tool would positively affect students' post-test performance.

Again we encountered the previously mentioned problems during our study. In this study too, students worked in groups, which created a *hierarchically nested dataset*. When we calculated the ICC of our dependent variable, post-test performance, we found an ICC of .32. This meant that 32% of the total variance was explained by group level variables and that the assumption of *independence* was violated. Finally, the variables we studied were measured at *different units of analysis*. Post-test performance, our dependent variable, was measured at the level of the student. In contrast, our independent variable, experimental condition (Graphical versus Textual Debate-tool) was measured at the level of the group, because each group was assigned to one of the two conditions. Finally, our analyses also included a covariate, pretest performance, which was again measured at the individual level.

If we chose to *ignore nonindepence* when analyzing the effects of the Graphical Debate-tool on students' post-test performance, we could use analysis of covariance (ANCOVA). The ANCOVA model would include post-test performance as the dependent measure of interest, condition (Graphical versus Textual Debate-tool) as the independent variable, and pre-test performance as a covariate. The results of this analysis can be found in Table 5. As can be seen in this Table, condition had a significant impact on post-test performance, $F(1, 82) = 3.98$, $p = .05$. In conclusion, if we adopt a strategy that ignores nonindependence, we would conclude that condition has a significant impact on post-test performance.

Table 5: Analysis of covariance for condition (Graphical versus Textual Debate-tool) on post-test performance.

	df	*MS*	*F*	η^2
Pretest performance (Covariate)	1	39.14	11.11**	.12
Condition (Graphical or Textual Debate)	1	14.07	3.98*	.05
Error	82	3.55		

* $p < .05$. ** $p < .01$.

Our second option would be to *aggregate* our data. This involves computing, for each group, the average post- and pre-test score of the individual group members. Using such a strategy, we find Graphical Debate groups to attain, on average, a post-test score of 13.02, while Textual Debate groups attain a an average score of 12.24. To test the effect of condition on post-test performance, we could again conduct an analysis of covariance, using post-test performance as the dependent variable, condition as the independent variable, and pre-test performance as a covariate. The results of this analysis are displayed in Table 6. As can be seen, the effects of condition are not significant if we adopt an aggregation strategy, $F(1, 37) = 0.28$, $p = .61$. This means we would conclude, in contrast to the previous strategy of ignoring nonindependence, the Graphical Debate-tool does not have a positive effect on students' post-test performance.

Table 6: Analysis of covariance for condition (Graphical versus Textual Debate-tool) on group level variables.

	df	*MS*	*F*	η^2
Pretest performance (Covariate)	1	11.39	6.27**	.28
Condition (Graphical or Textual Debate)	1	0.50	0.28	.02
Error	37	1.82		

** $p < .01$.

Our final option is to conduct a *multilevel* analysis. Our ML model then includes students' pre-test performance and condition (Graphical versus Textual Debate-tool). The results of this analysis can be found in Table 7. We found a significant effect of condition on

post-test performance, indicating that the Graphical Debate-tool helped students to perform better on the knowledge post-test, $\beta = 0.42$, SE = .22, $p = .03$.

Table 7: Multilevel analysis of the effect of condition (Graphical or Textual Debate-tool) on post-test performance.

	β	*SE*
Pre-test performance	0.28**	0.10
Condition (Graphical or Textual)	0.42*	0.22
Deviance	344.63	
Decrease in deviance	11.07**	

* $p < .05$. ** $p < .01$.

Table 8 summarizes the results of the different analysis strategies. As can be seen, the *p*-values are different if one strategy is chosen rather than another strategy. Especially when an aggregation strategy is chosen for the evaluation of the effect of the Graphical Debate-tool, a difference is noticeable. When we analyzed aggregated data we would draw a different conclusion – the Graphical Debate-tool does not affect post-test performance – compared to ignoring nonindependence or using MLA. This again highlights the importance of carefully using the appropriate data analysis strategy.

Table 8: Summary of differing results for effects of Graphical Debate-tool on students' post-test performance.

	Ignoring nonindependence	Aggregating data	Multilevel analysis
Statistical analysis	Analysis of covariance	Analysis of covariance	MLA
Significance of effect of condition	Significant, $p = .05$	Not significant, $p = .61$	Significant, $p = .03$
Conclusion	Positive effect of Graphical Debate-tool on post-test performance	No effect of Graphical Debate-tool on post-test performance.	Positive effect of Graphical Debate-tool on post-test performance

Example 3: Influence of representational guidance on essay quality

The study described in the previous example also provides the opportunity to highlight the effects of using a *disaggregation* strategy. Besides post-test performance, we also examined the effects of the Graphical Debate-tool on the quality of the essays written by groups. This effect was examined by measuring the number of topics covered in the essay and the quality of the essay. Because the essays were written by groups, this variable was a group-level measure: each group received one score for number of topics covered and quality of the essay. Thus, multilevel analysis is not necessary. However, in this could we could also have adopted a disaggregation strategy. This means each student within the group is given the same score for these two quality indicators. This leads to an increase of the sample size from 39 groups to 124 students. In Janssen et al. (in press), using *t*-tests, we found no significant differences with respect to the number of topics covered, $t = -0.55$, $p = .59$, and quality of the essay, $t = 2.00$, $p = .06$. However, when we disaggregate our data, and then use *t*-tests to examine the differences between the Graphical and Textual Debate tool, we find different *t*- and *p*-values, namely for number of topics covered, $t = -1.38$, $p = .17$, and for essay quality, $t = 3.24$, $p = .00$. This example shows that using a disaggregation strategy might lead to biased *t*- and *p*-values and even different conclusions (i.e., in the case of essay quality the conclusion would be different).

CONCLUSION AND DISCUSSION

In this chapter we discussed the data analytical problems CSCL researchers frequently encounter, namely hierarchically nested datasets, nonindependence of dependent variables, and differing units of analysis. We argued that, in order to take these problems into account,

MLA should be used. We also demonstrated that alternative analysis strategies such as ignoring nonindependence or aggregating or disaggregating data can lead to different results and possibly to mistakes regarding the significance or non-significance of these results. We therefore strongly advocate the use of MLA in CSCL research. Fortunately, more and more CSCL researchers are beginning to use this technique to answer their research questions.

It should be noted that we do not claim that in the cases where CSCL researchers used other analyses than MLA their conclusions are wrong. This need not be the case. However, these researchers do have an increased chance of committing Type I or Type II errors. We hope this chapter will contribute to an increased awareness of the risks of using traditional statistical techniques such as *t*-tests and ANOVAs, and future CSCL research will use MLA when this is appropriate.

Of course not all data-analytic problems that CSCL researchers encounter are solved by using MLA. Furthermore, MLA has its own limitations. First, MLA is mostly used when the dependent variable is measured at the interval level of measurement. Sometimes however, researchers may be interested in dichotomous (e.g., success or failure of group work) or categorical dependent variables (e.g., levels of knowledge construction). Although MLA techniques have been developed to incorporate these kinds of dependent variables (multilevel logistic regression, see Snijders & Bosker, 1999), they are rarely adapted to CSCL data.

Second, for an adequate analysis of collaborative learning using MLA, it is often suggested that large sample sizes at all levels (individual as well as group) are necessary (Cress, 2008; Maas & Hox, 2005). Maas and Hox, using a simulation study, demonstrated that only a small sample size at the group level (less than 50 groups) is problematic and leads to biased estimates. A small sample size at the individual level (groups consisting of five group members or less), does not appear to be problematic. This means that, in order to use MLA confidently for CSCL data, researchers should collect data about at least 50 groups.

CSCL researchers often employ less than 50 groups in their studies. Given the complexity of CSCL research and how time-consuming data collection and analysis often are, a sample size of at least 50 groups places a heavy burden on CSCL researchers.

Third, CSCL researchers are often interested in data over time. An example might be how familiarity with group members affects trust-development in CSCL environments over time. To investigate this question a researcher would collect data about trust levels on different occasions. This adds even more problems to analyzing CSCL data. The effects of familiarity on trust may not be the same at every measurement occasion (e.g., its effects may be greater at the beginning of the collaboration). Furthermore, the level of trust at measurement occasion 1 may also have an effect on the level of trust at occasion 2 (if trust was high at occasion 1, this may affect trust at occasion 2). This creates a new type of nonindependence: autocorrelation (Kenny et al., 2006). Again, MLA techniques have been developed to analyze time-series data (cf., Chiu & Khoo, 2003, 2005; Kenny et al., 2006), but they are not often used in CSCL research. CSCL researchers should therefore begin to investigate the possibilities of using MLA for time-series data.

Finally, MLA will not be a suitable technique to answer all research questions. Quite a lot CSCL research focuses on capturing the interactive processes that unfold between group members. In some cases researchers are interested in providing “thick” or “rich” descriptions of the collaborative process (Baker, 2003; Hmelo-Silver & Bromme, 2007). In such cases, MLA is obviously useless. Furthermore, it has been argued that studying intersubjective meaning making or group cognition should be the focus of CSCL research (Stahl, 2006; Suthers, 2006). This involves studying “how people make sense of situations and of each other” (Suthers, p. 321). Researchers with such a perspective on CSCL research could object to disentangling group and individual aspects of collaborative learning. They would argue that in order to understand the collaborative process, the group should be the unit of analysis,

not the individual. Again, if one has such an approach to studying CSCL, using MLA will not be a sensible strategy.

Fortunately, over the last years a large body of literature on MLA has been published. For the CSCL researcher who finds him- or herself faced with a hierarchically nested dataset and nonindependent observations, several good and accessible textbooks on the statistical and technical background of MLA are available (e.g., Hox, 2003; Snijders & Bosker, 1999). Furthermore, several good articles have been published about how to apply MLA to group and CSCL data (e.g., Bonito, 2002; Cress, 2008; Kenny et al., 2006; Kenny et al., 2002) and several CSCL articles have been published that can serve as an example (De Wever, Schellens, Valcke, & Van Keer, 2006; De Wever et al., 2007; Schellens et al., 2005; Schellens et al., 2007; Strijbos et al., 2004, 2007). Finally, several programs specifically designed for performing MLA are available, such as MLwiN (http://www.cmm.bristol.ac.uk/MLwiN/index.shtml) and HLM 6 (http://www.ssicentral.com/hlm/index.html). Moreover, the fact that conventional statistical software such as SPSS and SAS now incorporate procedures for carrying out MLA means that the possibility to perform MLA has become a possibility for many CSCL researchers.

CSCL research can still make progress by incorporating MLA in its repertoire of analysis techniques. It is an encouraging development that CSCL researchers are turning toward MLA more often. It is our hope and expectation that this development will continue and that CSCL researchers are going to find new ways to deal with the complex data analytical problems they are faced with. Ultimately, this will lead to a better understanding of how critical features of the CSCL-environment (e.g., support given by the environment), the group (e.g., composition), and the individual student (e.g., prior knowledge, motivation) affect social interaction and students' learning processes. Furthermore, when researchers

combine MLA with qualitative analyses in a mixed methods design (Leech & Onwuegbuzie, 2009) an even more complete picture of the CSCL process is possible.

REFERENCES

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 47-78). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bonito, J. A. (2002). The analysis of participation in small groups: Methodological and conceptual issues related to interdependence. *Small Group Research, 33*, 412-438.

Bonito, J. A., & Lambert, B. L. (2005). Information similarity as a moderator of the effect of gender on participation in small groups: A multilevel analysis. *Small Group Research, 36*, 139-165.

Bosker, R. J., & Snijders, T. A. (1990). Statistische aspecten van multi-niveau onderzoek [Statistical aspects of multilevel research]. *Tijdschrift voor Onderwijsresearch, 15*, 317-329.

Campbell, L., & Kashy, D. A. (2002). Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide. *Personal Relationships, 9*, 327-342.

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 analysis. *Small Group Research, 36*(5), 600-631.

Cress, U. (2008). The need for considering multilevel analysis in CSCL research: An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning, 3*, 69-84.

De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*, 57-85.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education, 46*, 6-28.

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. *Learning and Instruction, 17*, 436-447.

Dewiyanti, S., Brand-Gruwel, S., Jochems, W., & Broers, N. J. (2007). Students' experiences with collaborative learning in asynchronous Computer-Supported Collaborative Learning environments. *Computers in Human Behavior, 23*, 496-514.

Engelmann, T., Dehler, J., Bodemer, D., & Buder, J. (2009). Knowledge awareness in CSCL: A psychological perspective. *Computers In Human Behavior, 25*(4), 949-960.

Francescato, D., Porcelli, R., Mebane, M., Cuddetta, M., Klobas, J., & Renzi, P. (2006). Evaluation of the efficacy of collaborative learning in face-to-face and computer-supported university contexts. *Computers in Human Behavior, 22*, 163.

Guiller, J., & Durndell, A. (2007). Students' linguistic behaviour in online discussion groups: Does gender matter? *Computers in Human Behavior, 23*, 2240-2255.

Hmelo-Silver, C. E., & Bromme, R. (2007). Coding discussions and discussing coding: Research on collaborative learning in computer-supported environments. *Learning and Instruction, 17*, 460-464.

Hox, J. (2003). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Janssen, J., Erkens, G., & Kanselaar, G. (2007). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior, 23*, 1105-1125.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (in press). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science*.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451-477). Cambridge: Cambridge University Press.

Kenny, D. A. (1995). The effect of nonindependence on significance testing in dyadic research. *Personal Relationships, 2*, 67-75.

Kenny, D. A. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships, 13*, 279-294.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*, 422.

Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin, 119*, 138-148.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York/London: The Guilford Press.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83*, 126-137.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335-353.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity, 43*, 265-275.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *European Journal of Research Methods for the Behavioral and Social Sciences, 1*, 85-91.

O'Donnell, A. M., & O'Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review, 6*, 321-349.

Savicki, V., & Kelley, M. (2000). Computer mediated communication: Gender and group composition. *CyberPsychology & Behavior, 3*, 817-826.

Savicki, V., Kelley, M., & Lingenfelter, D. (1996). Gender and group composition in small task groups using computer-mediated communication. *Computers in Human Behavior, 12*, 209-224.

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.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

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

Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency: Using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 35*, 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.

Stylianou-Georgiou, A., Papanastasiou, E., & Puntambekar, S. (in press). Analyzing collaborative processes and learning from hypertext through hierarchical linear modelling. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methodologies, approaches and issues*.

Suthers, D. D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science, 7*, 254-277.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making. *International Journal of Computer Supported Collaborative Learning, 1*, 315-337.

Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences, 12*, 183-218.

Van der Meijden, H., & Veenman, S. (2005). Face-to-face versus computer-mediated communication in a primary school setting. *Computers in Human Behavior, 21*, 831-859.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner (Ed.), *Handbook of educational psychology* (pp. 841-873). New York: Simon & Schuster Macmillan.