Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups.

Bram De Wever, Hilde Van Keer, Tammy Schellens, and Martin Valcke

Department of Education, Faculty of Psychology and Educational Sciences, Ghent University, Belgium

Abstract

This study focuses on the process, output, and interpretation of *multilevel analyses* on *quantitative content analysis data* derived from asynchronous discussion group transcripts. The impact of role assignments on the level of knowledge construction reflected in students' contributions and the relation between message characteristics and these levels of knowledge construction is studied. Results show that summarisers' contributions and contributions focussing on theory, content moderating, or summaries result in significantly higher levels of knowledge construction. Multilevel modelling handles the hierarchical nesting, interdependency, and unit of analysis problem and is presented as a suitable technique for studying content analysis data from CSCL-environments.

1. Introduction

Within the field of computer-supported collaborative learning (CSCL), asynchronous discussion groups are often introduced as promising learning environments. The power of asynchronous text-based discussions lies in enhanced opportunities for students to interact with each other and in an increased time frame to reflect and search for additional information before contributing to the discussion (De Wever, Schellens, Valcke, & Van Keer, 2006; Pena-Shaff & Nicholls, 2004).

Researchers within this field are interested in the ongoing collaboration and the underlying interactive processes and more specifically in the impact of CSCL-environments on specific process and performance variables. However, analysing collaborative learning quantitatively is not a straightforward task, since the impact is affected by variables both at the level of the individual learner and the group. Individual learners are influenced by the social group and context to which they belong, and the properties of this group are in turn influenced by the individuals who make up that group (Hox & Maas, 2002). To take this into account, multilevel modelling techniques can be adopted to analyse the data at multiple levels. Although there is no general consensus regarding the appropriate statistical procedures to analyse data from content analysis of discussion transcripts, a number of methodological issues have been addressed by Chiu and Khoo (2005). Specifically, they support our view that multilevel analyses are an appropriate method to model content analysis data. These multilevel analysis techniques are highlighted in the present article. This article is not intended as a theoretical introduction to multilevel modelling, or as a complete overview of multilevel analysis approaches. However, this article presents a practical example of applying the analysis technique in the context of studying discussions groups.

The example fits in with the research tradition exploring the impact of different structuring approaches in online discussion tasks on the joint construction of knowledge. More specifically, the example builds on a study examining the effect of assigning roles to students on the knowledge construction processes in asynchronous discussion groups. To unravel students' knowledge construction, the discussion transcripts are analysed, as they contain information about both the group's collaborative process and the individuals' contributions and thus can serve as data for research (Meyer, 2004). In this respect, quantitative content analysis focussing on students' knowledge construction processes is performed to unlock the information captured in transcripts. This content analysis is based on the analysis model of Gunawardena, Lowe, and Anderson (1997) and is combined with multilevel modelling techniques in order to take the hierarchical structure of the data into account. Before starting a more elaborated discussion about this analysis technique, role assignment is discussed in short and presented as a critical scripting tool.

1.1. Roles as scripting tool

This study focuses on the impact of assigning roles as a scripting tool to support the process of social negotiation in asynchronous discussion groups. Roles compel students to focus upon their responsibilities in the discussion group and on the content of their contributions. Moreover, as roles are supposed to increase students' awareness of collaboration (Strijbos, Martens, Jochems, & Broers, 2004), we might expect students to collaborate better, resulting in higher levels of knowledge construction.

In the present study, the impact of the following roles has been studied: starter, theoretician, source searcher, moderator, and summariser. The starter is required to start off and give new impulses each time the discussion slacks off. The role of the moderator consists of monitoring the discussion, asking critical questions, and inquiring for others' opinions. The theoretician is asked to ensure that all relevant theoretical concepts were used in the discussion. The role of the source searcher comprises seeking external information on the discussion topics in order to stimulate other students to go beyond the scope of the course reader. The summariser is expected to post interim summaries during the discussion, focussing on identifying dissonance and harmony between the messages and drawing provisional conclusions, and a final summary at the end of the discussion.

The introduction of these roles is based on examples found in the literature, such as facilitator, resource person, summariser, starter, wrapper, discussion moderator, topic leader, and topic reviewer (Cohen, 1994; Hara, Bonk, & Angeli, 2000; Shotsberger, 1997; Tagg, 1994). On the other hand, the selection of the roles is based on the specific purpose of the discussion tasks, namely to stimulate students to actively discuss the content of the course manual and relevant external sources in order to get a grip on the different theoretical concepts introduced in the course. From an empirical point of view, earlier research already pointed at the positive impact of the role of a summariser in a discussion, resulting in significantly higher levels of knowledge construction (Schellens, Van Keer, & Valcke, 2005).

1.2. Content analysis

Neuendorf (2002, p. 10) defines content analysis as "a summarizing, quantitative analysis of messages that relies on the scientific method and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented". The aim of content analysis is to go beyond analyses based on counting the number of messages and to reveal information below the surface of the transcripts. In a previous article (De Wever et al., 2006), 15 different content analysis schemes were discussed in detail. In the present study, the interaction analysis model of Gunawardena et al. (1997) was applied. It focuses on the construction of knowledge through social negotiation, and distinguishes 5 levels, namely (a) sharing and comparing information, (b) identifying areas of disagreement, (c) negotiating meaning and co-construction of knowledge, (d) evaluation and modification of new schemas resulting from co-construction, and (e) reaching and stating agreement and application of co-constructed knowledge. This analysis scheme focuses on knowledge construction from a theoretical and empirical base (Marra, Moore, & Klimczak, 2004; Schellens & Valcke, 2005).

1.3. Multilevel modelling

The critical position of statistical analysis techniques has only recently been raised in CSCL-research. Within collaborative learning, learners are members of a group. The individual students and the social group can be “conceptualised as a hierarchical system of individuals and groups, with individuals and groups defined at separate levels of this hierarchical system” (Hox & Maas, 2002, p. 2). In this respect, the value of multilevel modelling is highlighted, since these models tackle problems that traditional unilevel statistical techniques are unable to cope with correctly.

In hierarchically structured settings, the assumption of independency for using the traditional analysis techniques is violated. With regard to the present study this means that data from students within a discussion group cannot be considered as completely independent because of the shared group history (Hox, 1994). In this respect, Hox and Maas (2002, p. 2) claim that “even if the analysis includes only variables at the lowest level, standard multivariate models are not appropriate. The hierarchical structure of the data creates problems, because the standard assumption of independent and identically distributed observations is generally not valid”. Due to the violation of the assumption of independence, conventional modelling can result in underestimation of standard errors. Researchers might reach conclusions about statistical significance and reject the null hypothesis because of small standard errors (Goldstein, 1995). In addition, even in situations where it is unlikely to make erroneous judgements, multilevel modelling provides more accurate estimates and should be used with data from natural groups, as “the existence of such data hierarchies is neither accidental nor ignorable” (Goldstein, 1995, p. 1).

Collaborative research designs entail that data are collected at different levels. They have to cope with the friction between individual-level versus group-level analysis (Flanagin, Park, & Seibold, 2004). Furthermore, cross-level interactions between variables on different levels of the hierarchy can influence outcome variables on a specific level. Because of the joint modelling of several variables at different levels, we encounter the methodological unit of analysis problem. By adopting multilevel modelling the hierarchical nesting, the interdependency, and the unit of analysis problem is handled in a more natural way, since this modelling approach is specifically geared to the statistical analysis of data with a clustered structure.

2. Method

2.1. Context

The study involved freshmen taking the course Educational Sciences at Ghent University. These students were randomly assigned to asynchronous discussion groups of 10 students. The discussion groups were an obligatory part of the course and were organised in addition to weekly face-to-face working sessions. The discussions were expected to foster students’ processing of the learning content and to promote discussion about the theoretical concepts presented in the face-to-face sessions and in the course manual. In the discussion groups, students were expected to solve authentic tasks. Taking into account the specific nature of discussing online, an introductory session was organised prior to the onset of the discussions, focussing on clarifying the aim of the discussions, the specific planning of different discussion themes, the different roles, the technical issues of the CSCL environment, and the evaluation criteria. All introductory information could be retrieved online. To ensure that students became familiar with the technical features of the online asynchronous discussion approach, a trial discussion session was organised during two weeks.

After the trial discussion, each group of students participated in four consecutive discussion themes. Each theme was organised during a three-week period. During this period, students collaborated independent of time and location. Participation in the discussions was obligatory and represented 25% of the final score. Students were required to contribute at least four messages per discussion theme. The four themes corresponded to four chapters in the course manual, namely behaviourism, cognitivism, constructivism, and evaluation. The authentic tasks in the discussion groups were identical for all groups.

2.2. Research design

Roles were introduced as a scripting tool during the first two discussions. Role-assignment was cut back after the second discussion theme, since it was expected that students would have interiorised the function of the roles. This transition from explicit role support to no role support is based on the assumption that fading of support should be an integral part of scaffolding, as outlined by Brown, Collis, and Duguid (1989).

At the start of the first discussion theme, five students were allocated the role of starter, moderator, theoretician, source searcher, and summariser respectively. These roles were assigned randomly and passed on to other students within the same group at the start of the second discussion theme. As stated before, none of the students were given roles during the last two discussion themes. At a general level, all students were encouraged to moderate, to summarise, and to add new discussion points, theory, and information. But students with a specific role were asked to do this in an explicit and regular way.

2.3. Data set and analysis instruments

The discussion transcripts of the 4 themes of 14 groups were randomly selected for content analysis. These transcripts were coded independently by four trained coders. The complete message was chosen as the unit of analysis. Complete messages are considered as the unit defined by the original author of the contributions, as suggested by Rourke, Anderson, Garrison, and Archer (2001).

Each contribution to a discussion reflects a specific level of social construction of knowledge. In order to determine these levels, the interaction analysis model of Gunawardena et al. (1997) was applied. The codes and descriptions of this model can be consulted in Table 3. Each message receives one code. This variable will serve as our dependent variable, indicating the degree of collaborative knowledge construction. When a message comprises elements of two different levels of knowledge construction, the highest level was assigned. For example, when a student shared information in order to argument why he or she disagrees with another student, this was coded as level 2 (disagreement) and not as level 1 (sharing new information).

Next to this content analysis scheme, an additional analysis model was developed to identify message characteristics along five different dimensions: moderating, summarising, adding new discussion points, adding theory, and adding external information. Indicators of different levels within these dimensions are presented in Table 1.

As opposed to the model of Gunawardena et al. (1997), the analysis scheme identifying message characteristics (ASIMeC) is specifically related to the different roles. Each unit of analysis is assigned a code along these dimensions. The scheme was developed to provide more information about the actual role adoption in the discussion groups. As to the dimension "adding external information" for example, the ASIMeC differentiates between "mentioning external sources" without linking the source to the ongoing discussion and "actively using and discussing new external sources".

Table 1
Analysis scheme identifying message characteristics (ASIMeC)

Dimension	Characteristic (code)	Description
Theory	No theory	Not referring to theoretical concepts
	Mentioning theory	Mentioning theoretical concepts
	Discussing theory	Actively using and discussing theoretical concepts
Source	No sources	Not referring to external sources
	Mentioning sources	Mentioning external sources
	Discussing sources	Actively using and discussing external sources
Summary	No summary	Not summarising information from other messages
	Minor summary	Summarising information from a number of messages
	Extensive summary	Summarising information of a substantial part of the discussion
Moderating	No moderating	No moderation tasks performed
	Organisational moderating	Organisational moderation tasks performed (e.g. planning)
	Content moderating	Moderation task as regards content performed (e.g. compare different statements, weigh up different messages)
	Organisational and content moderating	Combination of both moderation tasks
New points	No new points	No new points added to the discussion
	New points introduced	New points added to the discussion

2.4. Hypotheses

A first hypothesis focuses on the analysis of the transcripts of the first two discussions and explores the impact of the different *roles* on students' *level of knowledge construction through social negotiation*. More specifically, the level of knowledge construction of students adopting a role is compared with the level of knowledge construction of students without roles. It is hypothesised that students performing the role of starter, moderator, theoretician, source searcher, or summariser post messages reflecting higher levels of knowledge construction.

The second hypothesis focuses on the analysis of the four discussion themes and clusters two subhypotheses. First, we test whether a gradual increase in level of knowledge construction can be observed, since a learning effect could be expected in the course of the consecutive discussions. Secondly, this hypothesis concentrates on the relation between *message characteristics* and the *level of knowledge construction*. Messages reflecting characteristics such as summarising, moderating, introducing new discussion points, and debating theory and various sources are expected to reflect higher levels of knowledge construction as compared to contributions without these characteristics.

2.5. Statistical analysis

The data collected within the framework of the present study have a clear hierarchical structure. Every student belongs to one group. Furthermore, each message is written by one student. Therefore, multilevel modelling was applied. To test the first hypothesis, a three-level model was built. Messages are clustered within students that are nested within discussion groups. Taking into account that multilevel modelling is especially useful to analyse repeated measures (Snijders & Bosker, 1999), a specific type of hierarchical nesting was defined to test the second hypothesis: measurement occasions (in our case the four themes) nested within subjects (Hox, 1998). This results in a hierarchical structure in which messages are hierarchically nested within measurement occasions that are clustered within students who are in turn assigned to groups.

In view of testing both hypotheses, we start by calculating a random intercept null model. This model only contains an estimation of the intercept for the dependent variable, so there are no independent variables or predictors involved. In this null model, the total variance of students' level of knowledge construction is decomposed into between-group, between-students, and between-message variance. Next, explanatory variables are added to the models. Roles serve as a predictor for testing the first hypothesis. For the second hypothesis, both the measurement occasions (themes) and the dimensions of the ASIMeC serve as predictors. All models are discussed in detail in the results section.

The statistical package R 1.8.1. was used for the calculation of the interrater reliability coefficient Krippendorff's alpha. The descriptive results were calculated with SPSS 11.0.1. MLwiN 2.01. was used to perform the multilevel analysis. The multilevel models were estimated with the iterative generalised least squares (IGLS) procedure in order to build and compare the models. Restrictive iterative generalised least squares (RIGLS) was used to fine-tune the models. All reported parameter estimates are based on the RIGLS procedure. All analyses assume a 95% confidence interval. No centring was used.

2.6. Coding strategy and reliability

Four independent coders were trained during approximately 3 hours to perform the coding activity. They were introduced to the content analysis models, the underlying theoretical basis, and a number of examples to illustrate each coding scheme. After the training, transcripts were coded together for another 4 hours and the coding process was discussed and elaborated. Next, the transcripts were coded independently.

A number of transcripts was selected for calculating interrater reliability coefficients of the ASIMeC and the model of Gunawardena et al. (1997) (approximately 7% and 15% respectively). Table 2 presents the Krippendorff's alpha (α) interrater reliability coefficients. The values for Krippendorff's alpha were all situated within the classification 'fair to good agreement beyond chance'.

Table 2
Overview of the Krippendorff's alpha reliability coefficients

Variable	α
Level of knowledge construction (n=510)	0.53
Source (n=236)	0.75
Theory (n=236)	0.74
Summary (n=236)	0.62
Moderating (n=236)	0.59
New points (n=236)	0.63

3. Results

3.1. Descriptive results

In total 3345 messages were analysed (approximately 40,943 lines of text) with the interaction analysis model of Gunawardena et al. (1997). Table 3 presents an overview of the descriptive results.

Of these messages, approximately 2859 messages were analysed along the five dimensions of the ASIMeC. Table 4 gives an overview of these descriptive results.

Table 3
Overview of the codes based on the interaction analysis scheme of Gunawardena et al. (1997)

Level	Description	Frequency	Percent
1	sharing and comparing of information	2132	63.7
2	exploration of dissonance	658	19.7
3	negotiation of meaning	420	12.6
4	testing synthesis	95	2.8
5	agreement statements and applications	40	1.2

Table 4
Overview of the codes based on the ASIMeC

Dimension	Characteristic (code)	Frequency	Percent
Theory	No theory	828	29.0
	Theory mentioned	1357	47.5
	Theory discussed	671	23.5
Source	No sources	2526	88.4
	Source mentioned	168	5.9
	Source discussed	165	5.8
Summary	No summary	2697	94.3
	Minor summary	50	1.7
	Extensive summary	112	3.9
Moderating	No moderating	2264	79.2
	Organisational moderating	78	2.7
	Content moderating	506	17.7
	Organisational and content moderating	9	.3
New points	No new points	1816	63.5
	New points introduced	1042	36.5

3.2. Results for hypothesis 1

The null model shows that respectively 2.63%, 2.84%, and 94.53% of the total variance in students' level of knowledge construction is linked to differences between groups, between students within groups, and between students' messages. The group-level variance is not significantly different from zero ($\chi^2 = 3.415$, $df = 1$, $p = .065$), whereas the within-group between-student variance ($\chi^2 = 4.204$, $df = 1$, $p = .040$) and the variance between messages of students ($\chi^2 = 769.758$, $df = 1$, $p < .001$) are significantly different from zero.

In the final random intercepts model the five roles were contrasted with the reference category (no role). The estimates for this model are presented in Table 5. The intercept of 1.377 represents the mean level of knowledge construction for messages from students without roles. The mean level of knowledge construction reflected in messages from students with the role of starter, moderator, theoretician, and source searcher does not differ significantly from this mean. However, students with the role of summariser post messages with a significantly higher mean level of knowledge construction (mean = 1.377 + 0.321 = 1.698, $\chi^2 = 32.376$, $df = 1$, $p < .001$).

Table 5
Model estimates for the three-level analyses of levels of knowledge construction in students' messages

Parameter	Null model	Final Model
Fixed		
Intercept	1.416 (0.035)	1.377 (0.041)
Starter		-0.053 (0.058)
Moderator		0.057 (0.057)
Theoretician		0.051 (0.055)
Source searcher		-0.034 (0.056)
Summariser		0.321 (0.056)***
Random		
Level 3 – group		
σ^2_{v0}	0.012 (0.007)^{1}	0.013 (0.007)
Level 2 – student		
σ^2_{u0}	0.013 (0.006)^{*2}	0.010 (0.006)
Level 1 – message		
σ^2_{e0}	0.432 (0.016)^{***3}	0.426 (0.015)***

Values between brackets are standard errors
[1] 2.63%, [2] 2.84%, and [3] 94.53% of total variance
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

3.3. Results for hypothesis 2

Firstly, a four-level random intercepts null model was estimated, with messages (level 1) hierarchically nested within the themes (measurement occasions, level 2) that are clustered within students (level 3) assigned to groups (level 4). This model is presented as model A in Table 6.

As can be seen in the random part of this model, the variances the on group, theme, and messages level are significantly different from zero: 2% of the total variance in students' levels of knowledge construction is situated at the group level ($\chi^2 = 4.274$, $df = 1$, $p = .039$), 7% is situated at the theme level (measurements occasions) ($\chi^2 = 25.951$, $df = 1$, $p < .001$), and 91% of the variance arises from differences between messages within measurement occasions ($\chi^2 = 1440.268$, $df = 1$, $p < .001$). No part of the total variance can be assigned to the level of the individual students.

Secondly, a compound symmetry model (model B) was estimated. This is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999). This model allows us to explore whether a learning effect occurs throughout the successive themes. The differences between the themes are explicitly modelled by contrasting theme 2, theme 3, and theme 4 with the reference category (theme 1). This model achieves a better fit than the four-level null model, for the difference in deviance of both models – which can be used as a test statistic having a chi-squared distribution, with the difference in number of parameters as degrees of freedom (Snijders & Bosker, 1999) – is highly significant ($\chi^2 = 145.036$, $df = 3$, $p < .001$).

The intercept of 1.505 is to be considered as the overall mean level of knowledge construction in theme 1 across all messages, students, and groups. As presented in Table 6 (model B), the mean level of knowledge construction in theme 4 does not differ significantly from this intercept (mean = 1.505 + 0.033 = 1.538, $\chi^2 = 0.605$, $df = 1$, $p = .437$). However, messages in theme 2 reflect a significantly lower level of knowledge construction (mean = 1.505 - 0.169 = 1.336, $\chi^2 = 15.738$, $df = 1$, $p < .001$), while messages in theme 3 reflect a significantly higher level of knowledge construction (mean = 1.505 + 0.393 = 1.898, $\chi^2 = 94.939$, $df = 1$, $p < .001$).

Table 6
Model estimates for the four-level analyses of levels of knowledge construction.

Parameter	Model A	Model B	Model C
Fixed			
Intercept	1.573 (0.040)	1.505 (0.048)	1.321 (0.058)
Theme 2 (cognitivism)		-0.169*** (0.043)	-0.176*** (0.047)
Theme 3 (constructivism)		0.393*** (0.040)	0.433*** (0.044)
Theme 4 (evaluation)		0.033 (0.043)	0.060 (0.047)
Theory mentioned			0.223*** (0.038)
Theory discussed			0.238*** (0.048)
Source mentioned			-0.273*** (0.067)
Source discussed			0.061 (0.069)
Minor summary			0.336** (0.122)
Extensive summary			0.864*** (0.083)
Organisational moderating			-0.399*** (0.099)
Content moderating			0.161*** (0.042)
Organisational and content moderating			-0.463 (0.281)
New points			-0.074* (0.037)
Random			
Level 4 – group			
σ^2_{f0}	0.017* (0.008)	0.020* (0.009)	0.020* (0.010)
Level 3 – student			
σ^2_{v0}	0.000 (0.000)	0.008 (0.005)	0.000 (0.000)
Level 2 – theme			
σ^2_{u0}	0.057*** (0.011)	0.000 (0.000)	0.007 (0.008)
Level 1 – message			
σ^2_{e0}	0.733*** (0.019)	0.739*** (0.018)	0.688*** (0.019)
Model fit			
Deviance	8675.480	8530.444	7066.443
χ^2		145.036	1464.001
df		3	10
p		<.001	<.001
Reference		Model A	Model B

Values between brackets are standard errors
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

The contradictory findings that messages in theme 2 reflect a significantly lower and messages in theme 3 reflect a significantly higher level of knowledge construction, were briefly explored by taking task complexity into account. To verify the impact of the different discussion themes' complexity, all participating students were asked to rate the difficulty of each assignment. Task complexity increased significantly from the first to the three subsequent themes, and the second assignment was identified as the most complex. These self-reported complexity rates were included in the analysis model, revealing no significant differences in levels of knowledge construction between the consecutive themes, except for a significantly higher level in the third discussion theme. Although these results are interesting, they are not discussed in detail, as this post hoc analysis was not the scope of this study.

In a third model (model C), the five dimensions of the ASIMeC are added as extra explanatory variables. This results in a significantly better fit of the model ($\chi^2 = 1464.001$, *df* = 10, $p < .001$). In this model, 11.4% of the variance in levels of knowledge construction between the messages (R_1^2, see Snijders & Bosker, 1999) is explained by the characteristics of the messages. The intercept of 1.321 in this model represents the mean level of knowledge construction for messages in theme 1 that do not include theory, sources, summaries, moderation issues, or new points. Parallel to the compound symmetry model (model B), the mean level of knowledge construction of messages in theme 4 does not differ significantly from this intercept (mean = 1.321 + 0.060 = 1.381, $\chi^2 = 1.679$, *df* = 1, $p = .195$), messages in theme 2 reflect a significantly lower level of knowledge construction (mean = 1.321 - 0.176 = 1.145, $\chi^2 = 14.189$, *df* = 1, $p < .001$), and messages in theme 3 reflect a significantly higher level of knowledge construction (mean = 1.321 + 0.433 = 1.754, $\chi^2 = 95.256$, *df* = 1, $p < .001$).

Concerning *theory*, both mentioning and discussing theory leads to a significantly higher mean level of knowledge construction (mean = 1.321 + 0.223 = 1.544, $\chi^2 = 33.950$, *df* = 1, $p < .001$ and mean = 1.321 + 0.238 = 1.559, $\chi^2 = 24.764$, *df* = 1, $p < .001$ respectively). The same goes for the variable *summary*: both minor summaries and extended summaries lead to a significantly higher mean level of knowledge construction (mean = 1.321 + 0.336 = 1.657, $\chi^2 = 7.566$, *df* = 1, $p = .006$ and mean = 1.321 + 0.864 = 2.185, $\chi^2 = 108.537$, *df* = 1, $p < .001$ respectively).

With regard to *sources*, the results are somewhat different: messages in which students mention but do not discuss new sources reflect a significantly lower mean level of knowledge construction, whereas messages including this discussion of the external sources do not reflect a significant different level of knowledge construction (mean = 1.321 - 0.273 = 1.048, $\chi^2 = 16.510$, *df* = 1, $p < .001$ and mean = 1.321 + 0.061 = 1.382, $\chi^2 = 0.779$, *df* = 1, $p = .377$ respectively).

Next, as to *moderating*, the mean level of knowledge construction is significantly lower for messages containing organisational moderation (mean = 1.321 - 0.399 = 0.922, $\chi^2 = 16.378$, *df* = 1, $p < .001$), significantly higher for messages comprising content moderation (mean = 1.321 + 0.161 = 1.482, $\chi^2 = 14.803$, *df* = 1, $p < .001$), and not significantly deviant for messages containing both organisational and content moderating (mean = 1.321 - 0.463 = 0.858, $\chi^2 = 2.724$, *df* = 1, $p = .099$).

Additionally, messages introducing *new points* reflect a significantly lower mean level of knowledge construction (mean = 1.321 - 0.074 = 1.274, $\chi^2 = 4.038$, *df* = 1, $p = .044$).

In Table 7 we compared the results of the four-level analysis with a unilevel OLS regression analysis. Although some parameters are quite alike, others show some differences. This results in differences in the p-values, especially when parameters are not highly significant. Compared to the p-values of the four-level model, the p-values of the unilevel regression vary from 1.60 times smaller to 1.76 times larger.

Table 7
Comparison of the parameter estimates and p-values of the four-level analysis and unilevel regression analysis of the data in Model C.

	Model C (four-level model from Table 6)			Model C (unilevel OLS model)		
	Estimate	s.e.	p-value	Estimate	s.e.	p-value
Fixed						
Intercept	1.321	(0.058)	< .001	1.309	0.041	< .001
Theme 2 (cognitivism)	-0.176	(0.047)	< .001	-0.173	0.046	< .001
Theme 3 (constructivism)	0.433	(0.044)	< .001	0.443	0.043	< .001
Theme 4 (evaluation)	0.060	(0.047)	.195	0.066	0.045	.146
Theory mentioned	0.223	(0.038)	< .001	0.240	0.039	< .001
Theory discussed	0.238	(0.048)	< .001	0.256	0.048	< .001
Source mentioned	-0.273	(0.067)	< .001	-0.308	0.068	< .001
Source discussed	0.061	(0.069)	.377	0.078	0.070	.265
Minor summary	0.336	(0.122)	.006	0.323	0.123	.009
Extensive summary	0.864	(0.083)	< .001	0.833	0.083	< .001
Organisational moderating	-0.399	(0.099)	< .001	-0.437	0.099	< .001
Content moderating	0.161	(0.042)	< .001	0.155	0.042	< .001
Organisational and content moderating	-0.463	(0.281)	.099	-0.399	0.283	.158
New points	-0.074	(0.037)	.044	-0.82	0.037	.025
Random						
Level 4 – group						
σ^2_{f0}	0.020	(0.010)	.034			
Level 3 – student						
σ^2_{v0}	0.000	(0.000)				
Level 2 – theme						
σ^2_{u0}	0.007	(0.008)	.341			
Level 1 – message						
σ^2_{e0}	0.688	(0.019)	< .001	0.714	0.019	< .001
Model fit						
Deviance			7066.443			7121.155

Values between brackets are standard errors

4. Discussion

A first conclusion that can be drawn from the results for hypothesis 1 is that multilevel modelling is an appropriate technique to analyse content analysis data, as the between-students and between-messages variance is significantly different from zero. The large proportion of variance situated at the level of the messages indicates that a student's messages generally reflect a whole range of different levels of knowledge construction, while only rather small differences between students and between groups can be observed.

Secondly, with respect to the impact of role assignment on the level of knowledge construction in students' messages, we can conclude that only the role of summariser has a significantly positive effect. The other roles do not result in significantly higher levels of knowledge construction. This finding confirms previous research (Schellens et al., 2005) that studied the influence of four different roles (theoretician, source searcher, moderator, and summariser) and revealed that only students who perform the role of the summariser submit messages that reflect significantly higher levels of knowledge construction.

In this respect, it can be concluded that although the introduction of roles seems to increase students' awareness of group interaction and collaboration (Strijbos et al., 2004), this does not necessarily lead to an increase in students' knowledge construction. The positive effect of the summariser can be attributed to the fact that this student is expected to post interim summaries during the discussions, and this requires him/her to identify similarities or differences between the messages, to develop a general overview, to consider all parties and opinions, etc. These extra activities clearly push higher levels of knowledge construction.

However, with the exception of the role of the starter, also the other roles might require this type of higher level activities. Yet, considering the analysis results, this does not seem to be the case. The differential impact of the roles might be due to the fact that the task of the summariser is more explicit, more transparent, and more concrete for the students. In this respect, further research is needed to clarify this differential impact and to get a better understanding of role interpretation, adoption, and execution. Furthermore, next to focussing on the contribution of students performing roles, it might also be interesting to concentrate on other students' contributions, especially those following on role-related messages.

As to the second hypothesis, the results again reveal the importance of multilevel modelling, since the between-groups, between-themes, and between-messages variance is significantly different from zero. Again, the largest proportion of variance is situated at the message level, pointing towards large variability in levels of knowledge construction between student messages. Furthermore, as a learning effect could occur when students get acquainted with the CSCL-environment and master the necessary discussion skills, a gradual increase of students' level of knowledge construction throughout the different discussion themes was expected. However, the results do not completely support this assumption. The findings more specifically reveal that with reference to the first discussion theme, contributions reflect significantly lower levels of knowledge construction in theme 2, significantly higher levels in theme 3, and no significantly deviant levels of knowledge construction in theme 4. In this respect, it can be argued that the level of knowledge construction attained by students does not only depend on the increase in experience and discussion skills. Furthermore, given the decrease in knowledge construction in theme 2, and the increase in theme 3, it is unlikely that these effects can be attributed to the fact that roles were no longer assigned to the students. Other factors also appear to be important. In this respect, Schellens, Van Keer, Valcke, and De Wever (in press) refer to the significance of task characteristics. More specifically, the impact of task complexity appears to be important: when tasks are too complex, students' levels of knowledge construction are also significantly lower. On the other hand, when the assignments are overly straightforward, it is expected that students are hardly challenged and that the quality of the contributions drops. Based on the finding that, except for a significantly higher level in the third discussion, no significant differences in levels of knowledge construction are revealed when adding complexity to the model, it can be argued that the significant decrease in levels of knowledge construction in the second discussion theme can be attributed to a perceived high level of complexity in this assignment. This finding points at the importance of well-considered task design to foster knowledge construction.

In addition to the hypothesis of a growing trend in knowledge construction throughout the successive discussion themes, higher levels of knowledge construction were also expected for contributions mentioning and discussing theory, mentioning or discussing new sources, including minor and extensive summaries, containing organisational and/or content moderating, and introducing new points. The results corroborate this hypothesis for both mentioning and discussing theory, posting minor and extensive summaries, and introducing contributions including content moderating.

In combination with the frequent occurrence of theory, the significant effects of mentioning and discussing theory confirm that discussing theoretical concepts is probably an essential factor influencing knowledge construction.

Regarding summaries, both contributing minor and extensive summaries results in higher levels of knowledge construction.

As to moderating, contributions including organisational moderating reflect significantly lower levels of knowledge construction, whereas contributions including content moderating reflect significantly higher levels of knowledge construction. These results can be attributed to the fact that the former contributions focus on planning and organisation of the discussions and, as such, do not actually influence knowledge construction. In contrast, messages including content related moderation invoke knowledge construction activities. Nevertheless, organisational moderating might be a prerequisite for knowledge construction as it is important to guide the discussion process.

With respect to discussing new sources, no significant positive impact on knowledge construction is observed. On the contrary, merely mentioning sources, without explicitly discussing them and linking them to the ongoing discussions, even leads to significantly lower levels of knowledge construction. This is in line with the aforementioned research of Schellens et al. (2005).

Finally, contributions comprising new points result in lower levels of knowledge construction.

The differences between the OLS and the multilevel results show the importance of taking the hierarchical nesting into account for some parameters. Schreiber and Griffin (2004) argue that the OLS standard errors can be underestimated due to the fact that the variances at higher levels are not taken into account. This might have consequences for hypothesis testing. In the present study, the standard errors, and most parameters, are quite similar and the results of both procedures lead us to the same conclusions. However, as can be seen in Table 7, some parameters are estimated differently. As a result, the p-values fluctuate, which might have led to erroneous conclusions. Goldstein (1995) for, instance presented an example of a study in which OLS results were statistically significant but where the significant differences disappeared when multilevel analyses were performed, taking the groupings of the subjects into account.

5. Conclusion

Critical questions about the choice of statistical analysis techniques to study quantitative content analysis data have only recently been raised in the CSCL literature. Within this context, the present article focuses on the potential of adopting multilevel modelling methodologies. One of the main reasons for applying multilevel analysis is the fact that the use of unilevel analysis methods on multilevel data can have baleful consequences. Since multilevel modelling handles the hierarchical nesting, the interdependency, and the unit of analysis problem in a natural way, it is a suitable technique for analysing the interaction in collaborative learning environments in general, or content analysis data from CSCL-environments in particular. Especially for research in ecologically valid settings studying natural groups, multilevel modelling is a worthwhile alternative for traditional analysis techniques. Although a demonstration of the full power of multilevel modelling was beyond the scope of this article, the process, output, and interpretation of the specific analyses of this study have been described in detail. The results reveal that applying multilevel models on content analysis data can be used to gain an in depth understanding of the nature of collaborative learning processes.

Acknowledgment

This study was funded by BOF2002/DRMAN/013 of Ghent University.

References

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher, 18,* 32-42.

Chiu, M. M. & Khoo, L. (2005). A New Method for Analyzing Sequential Processes: Dynamic Multilevel Analysis. *Small Group Research, 36,* 600-631.

Cohen, E. G. (1994). *Designing Groupwork. Strategies for the Heterogeneous Classroom.* (2nd ed.) New York: Teachers College Press.

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.

Flanagin, A. J., Park, H. S., & Seibold, D. R. (2004). Group performance and collaborative technology: A longitudinal and multilevel analysis of information quality, contribution equity, and members' satisfaction in computer-mediated groups. *Communication Monographs, 71,* 352-372.

Goldstein, H. (1995). *Multilevel statistical models*. (2nd ed.) London: Edward Arnold.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17,* 397-431.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28,* 115-152.

Hox, J. J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research, 22,* 300-318.

Hox, J. J. (1998). Multilevel modeling: When and why. In R. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147-154). New York: Springer-Verlag.

Hox, J. J. & Maas, C. J. M. (2002). Sample sizes for multilevel modeling. In J.Blasius, J. J. Hox, E. De Leeuw, & P. Schmidt (Eds.), *Social Science Methodology in the New Millennium. Proceedings of the Fifth International Conference on Logic and Methodology. Second expanded edition.* Retrieved April 1, 2006, from http://www.fss.uu.nl/ms/jh/publist/simnorm1.pdf

Marra, R. M., Moore, J. L., & Klimczak, A. K. (2004). Content analysis of online discussion forums: a comparative analysis of protocols. *Educational Technology Research and Development, 52,* 23-40.

Meyer, K. A. (2004). Evaluating online discussions: Four different frames of analysis. *Journal of Asynchronous Learning Networks, 8*. Retrieved April 1, 2006, from http://www.sloan-c.org/publications/jaln/v8n2/pdf/v8n2_meyer.pdf

Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.

Pena-Shaff, J. B. & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education, 42,* 243-265.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological Issues in the Content Analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12,* 8-22.

Schellens, T. & Valcke, M. (2005). Collaborative learning in asynchronous discussion groups: What about the impact on cognitive processing? *Computers in Human Behavior, 21,* 957-975.

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. (in press). Learning in asynchronous discussion groups: A multilevel approach to study the influence of student, group, and task characteristics. *Behaviour & Information Technology*.

Schreiber, J. B. & Griffin, B. W. (2004). Review of multilevel modeling and multilevel studies in the Journal of Educational Research (1992-2002). *Journal of Educational Research, 98,* 24-34.

Shotsberger, P. G. (1997). Emerging Roles for Instructors and Learners in the Web-Based Instruction Classroom. In B.H.Khan (Ed.), *Web-based Instruction* (pp. 101-106). New Jersey: Educational Technology Publications.

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

Strijbos, J. W., Martens, R. L., Jochems, W., & Broers, N. J. (2004). The effect of functional roles on group efficiency: Using multilevel modelling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 35,* 195-229.

Tagg, A. C. (1994). Leadership from Within: Student Moderation of Computer Conferences. *The American Journal of Distance Education, 8,* 40-50.