Content analysis: What are they talking about?

Jan Willem Strijbos, Rob Martens, Frans Prins, Wim Jochems

CONTENT ANALYSIS: WHAT ARE THEY TALKING ABOUT?

doi:{10.1016/j.compedu.2005.04.002 }

Please refer this manuscript as:

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers and Education*, *46*, 29-48.

Running head: CONTENT ANALYSIS: WHAT ARE THEY TALKING ABOUT?

Content analysis: What are they talking about?

Jan-Willem Strijbos[1], Rob L. Martens, Frans J. Prins, & Wim M. G. Jochems

Open University of the Netherlands

[1] Correspondence can be sent to Jan-Willem Strijbos, Open University of the Netherlands, Educational Technology Expertise Center, P.O. Box 2960, 6401 DL, Heerlen, The Netherlands. E-mail: jan-willem.strijbos@ou.nl

Abstract

Quantitative content analysis is increasingly used to surpass surface level analyses in Computer-Supported Collaborative Learning (e.g., counting messages), but critical reflection on accepted practice has generally not been reported. A review of CSCL conference proceedings revealed a general vagueness in definitions of units of analysis. In general, arguments for choosing a unit were lacking and decisions made while developing the content analysis procedures were not made explicit. In this article, it will be illustrated that the currently accepted practices concerning the 'unit of meaning' are not generally applicable to quantitative content analysis of electronic communication. Such analysis is affected by 'unit boundary overlap' and contextual constraints having to do with the technology used. The analysis of e-mail communication required a different unit of analysis and segmentation procedure. This procedure proved to be reliable, and the subsequent coding of these units for quantitative analysis yielded satisfactory reliabilities. These findings have implications and recommendations for current content analysis practice in CSCL research.

Keywords: content analysis, methodology, reliability, unit of analysis, segmentation, coding

Content analysis: What are they talking about?

Koschmann (1996) identified computer-supported collaborative learning (CSCL) as an emerging field in educational research. Considerable attention has been paid to theoretical debate, as well as to technical and pedagogical support of collaborative learning. In comparison, however, less attention has been paid to issues of methodology and analysis methods (Strijbos, Kirschner, & Martens, 2004).

Initially, analyses in CSCL and computer-mediated communication (CMC) research focused on questionnaires or surface level characteristics of the communication (Harasim, Hiltz, Teles, & Turoff, 1995). For example, participation degree was determined by the number of messages sent (Harasim, 1993), and it was assumed that the mean number of words in a message was positively related to the quality of that message's content (Benbunan-Fich & Hiltz, 1999). Surface level measurements are still used and several methods have been added such as 'thread-length' (Hewitt, 2003) and 'social network analysis' (SNA; Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003). However, it is now widely acknowledged that surface level methods provide at best a rough analysis of the communication. Furthermore, the quality of group performance (product or grade) provides no insight into the actual collaborative process and contextual factors that affect collaboration. Hence, it is imperative that the communication is subjected to content analysis to determine *why* one student contributes more or appears to be more influential in a group (Strijbos, Martens, Jochems, & Broers, 2004).

Analysis of communication transcripts has gained attention in the past decade (Gunawardena, Lowe, & Anderson, 1997; Hara, Bonk, & Angeli, 2000; De Laat & Lally, 2003; Wang, Laffey, & Poole, 2001) and roughly two approaches can be derived from the CSCL literature. In the 'quantitative' approach the communication is coded, summarised and frequencies/percentages

are used for comparisons and/or statistical testing. This approach contrasts with the 'qualitative' view, that uses methods such as participant observation (Louca, Druin, Hammer, & Dreher, 2003), case summaries (Lally & De Laat, 2003) and ethnomethodology (Stahl, 2002b) to infer trends or a specific phenomenon in transcripts (Miles & Huberman, 1994) – without computing frequencies for statistical testing. This distinction also coincides with that of a prospective and retrospective analysis orientation (Strijbos, Martens, & Jochems, 2004).

The statistical comparisons in the quantitative approach require a hypothesis – derived from theory – formulated in advance (prospective), whereas the aim of 'understanding' a phenomenon (retrospective) requires less explicit a priori expectations or even none (e.g., in a grounded theory approach). Reliability is a concern in both approaches, but treated differently. In the quantitative approach, reliability is expressed in a numeric value indicating the level of agreement between two independent coders. In the qualitative approach reliability (credibility) is established through using multiple analysts, comparing two or more interpretive perspectives of independent coders and/or triangulation with external sources or quantitative data (Elliot, Fischer, & Rennie, 1999; Madill, Jordan, & Shirley, 2000).

In this article the importance of reliability for quantitative content analysis will be illustrated with an account of the development of a segmentation procedure and its impact on the coding of the communication. It will be argued that the quantitative approach – specifically its application in CSCL research – requires more rigour regarding reliability to warrant the apparent 'accuracy' of conclusions. Lack of reliability increases the probability of Type II errors (wrongly accepting the null-hypothesis) and, to a smaller degree, Type I errors (wrongly rejecting) occur.

Conclusions derived from statistical tests of data, for which the reliability of the method used is 'unknown', are questionable and should be treated with caution. Furthermore, reliability does

not only apply to 'assigning codes', but in those instances where the granularity of the unit of analysis is very small, reliability also applies to determining those 'units'. Rourke, Anderson, Garrison and Archer (2001) conclude: "Characteristics such as objectivity and reliability are not accidental features of some studies: rather, they are important criteria for any studies using this technique." (p. 20). Neuendorf (2002) even states that "Without the establishment of reliability, content analyses measures are useless." (p. 141).

To illustrate why reliability is important, consider how 'questionnaires' should be treated methodologically speaking. If a questionnaire is used, an alpha statistic should be reported to warrant the internal consistency of items measuring the psychological construct. If a previously constructed questionnaire is used, two statistics should be reported: the original alpha, as well as the alpha pertaining to the current study. If the questionnaire is adapted, again, the original alpha, as well as the alpha of the adapted questionnaire should be reported. Unfortunately, statistical tests on 'quantitative content analysis data with an unknown reliability' are still reported (see Pata & Sarapuu, 2003; Rasku-Puttonen, Eteläpelto, Arvaja, & Häkkinen, 2003).

The remainder of this article is structured as follows: first, the need for a clear definition of the 'unit of analysis' and the variety in 'units' used is illustrated through a review of papers in recent CSCL conference proceedings. The next section provides information about the research project for which a content analysis procedure was developed, followed by a section discussing the original procedure: segmentation rules and coding categories and rules. In the fourth section four factors will be discussed that – in retrospect – affect the applicability of a 'unit of analysis'. It will be shown that problems with reliably determining the unit of analysis turned out to be the primary cause for the failure of the original procedure. In the next section an alternative 'unit of analysis' is introduced, as well as a procedure to segment the communication in these units. In

addition, the reliability of this segmentation procedure (and its computation), revision of coding categories, and the reliability of this coding scheme, will be elaborated. Finally, the implications and recommendations for content analysis methodology in CSCL research will be discussed.

Defining and determining the unit of analysis: a review

Rourke et al. (2001) distinguish five types of units. From large to small they are a message (e-mail or forum contribution), paragraph (section), 'unit of meaning' (or thematic unit), sentence (or syntactical unit) and illocution. The most frequently reported units are a message, a 'unit of meaning' and the sentence. The definition of a unit of analysis, however, is often vague, which makes it hard to distinguish between them. Veldhuis-Diermanse (2002), for example, defines a 'unit of meaning' – following Henri (1992) and Chi (1997) – as a unit that represents "*an idea*, argument chain or discussion topic" (p. 46), whereas Muukkonen, Lakkala and Hakkarainen (2001) define the 'proposition unit' as "representing *a single idea*" (p. 462). These definitions illustrate the lack of clarity: a 'unit of meaning' and 'proposition' are both defined as 'an idea'. Moreover, the argumentation for choosing a specific 'unit of analysis' is rarely provided.

To assess the current state of the art with respect to units used in CSCL research, the CSCL conference proceedings (2001, 2002 and 2003) were reviewed. A conference is the primary forum where innovative methods are discussed and reflection on current practice is stimulated. Also, journal publications take considerably more time to appear, thus a conference proceeding reveals new developments faster. The review included 14 out of 18 papers in the 'analysis track' of CSCL 2001 (Dillenbourg, Eurelings, & Hakkarainen, 2001), 2 out of 5 long 'methodology' and 4 out of 30 short 'qualitative analyses' papers presented at CSCL 2002 (Stahl, 2002a), and 11 out of all 60 papers in the CSCL 2003 proceedings (Wasson, Ludvigsen, & Hoppe, 2003). Table 1 presents an overview of the various unit types reported (see Appendix A for an overview

of the studies and their unit of analysis).

Insert Table 1 about here

This review reveals that in 9 out of 31 (.29) papers the unit is not indicated and/or defined. Although conference papers in general do not offer the opportunity to describe the research in full, the content analysis methodology used should be clearly described or at least indicated – even if the paper length is restricted (also note that half of the studies in the Rourke et al. review do not report any reliability statistic). Moreover, none of the 13 out of 31 (.42) papers using a units smaller than a message report an intercoder reliability statistic for segmentation in units. In addition, the segmentation procedure is either non-existent or is not described. In contrast, all of the papers provide intercoder reliability for assigning codes.

The review by Rourke et al. (2001), in which nineteen studies (conducted between 1991–2000) of 'asynchronous text-based quantitative content analysis' were reviewed, supports this picture. Only ten out of nineteen studies reviewed report a reliability statistic (either proportion agreement or Cohen's kappa) and it is unclear whether the statistic refers to the 'unit of analysis', 'coding' or a combination. An exception is the recent study by Fischer, Bruhn, Gräsel and Mandl (2002) who define their unit as 'speech acts' and segmentation reliability is addressed but limited to the remark that transcripts were 'segmented by trained evaluators' (p. 220).

In sum, most conference papers and CSCL articles provide information on the reliability of their coding categories but little information is provided on segmentation. In addition, virtually no information is given on the process involved in developing a content analysis procedure (one exception is Chi (1997) on face-to-face communication), nor is argumentation provided for the

decisions made during the construction of the whole coding procedure. Most decisions appear to have been made intuitively or based on accepted practices. In the next sections the process of developing a content analysis procedure will be discussed.

Developing a content analysis procedure: original approach

The project for which a content analysis procedure was to be developed, is set in higher/distance education in the domain of 'policy development'. Students collaborate in groups of four. They have to collaboratively write a policy report containing a recommendation on reorganisation of local administration. They communicate only via e-mail. In such asynchronous CSCL settings, where group members are not present at the same time and place, coordination conflicts are very likely to occur (Benbunan-Fich & Hiltz, 1999). Clearly, some type of support should be provided to overcome coordination difficulties. On approach is the use of pedagogical support or so-called scripts (see Weinberger & Fischer, this issue), such as using roles. Roles can facilitate cohesion and responsibility, which indirectly affect coordination (Mudrack & Farrell, 1995). Roles can be defined as functions and/or responsibilities that guide behaviour and coordinate group interaction (Hare, 1994). In order to support coordination, functional roles were implemented in half of the groups in this course. The hypothesis was that these roles decrease the amount of coordinative statements ('who-does-what-when') in favour of statements that focus on the 'content' (solving the assignment) and thus affect group performance due to an increase in collaboration efficiency. All e-mail communication was to be coded and quantified for statistical comparison of both research conditions.

Procedure

The original procedure was developed at a time that the data collection of the project was still in progress, so it was decided to test the procedure on a similar dataset: six students collaborating

via e-mail in the domain of 'corporate law'. In this stage the choices were guided by the accepted practices and the 'unit of meaning' was applied (Aviv, 1999; Hara, Bonk, & Angeli, 2000; Henri, 1992; Newman, Webb, & Cochrane, 1995). Although proponents of a 'unit of meaning' argue that segmentation and coding should be performed simultaneously, it was decided to separate segmentation and coding to be more precise. The 'meaningfulness' of a statement should not be determined by coding categories (or a researcher). The statement "I went to the beach." may not have meaning with respect to the coding scheme, but this does not automatically reduce it to having no meaning. If segmentation and coding are combined, then all the instances where two independent coders disagree should still be carefully documented and the unit (or part) should be recoded by both coders in order to compute a reliability statistic (Prins, Busato, Elshout, & Hamaker, 1998). The original procedure included five rules to guide the segmentation in units of meaning: an e-mail message is a fixed unit and the message order is ignored, the salutation and close will be ignored, the unit of analysis is the 'unit of meaning', a post-script is considered to be a separate 'unit of meaning', and in case of a summation all points sharing a single 'meaning' are considered one 'unit of meaning'.

In a similar vein four rules were developed to aid the coding process: in case activities are organised according to the content of the task (e.g., splitting the task) these units are regarded as 'coordinative' and not as 'content', if a single unit can be assigned more than one category, the unit is split, in case a single unit can be assigned more than one category – and it can be split – it is allowed to assign two codes, and if two consecutive units are assigned the same code they will remain separate units.

Subsequently each unit was assigned a single code (italicised) from each of four main categories:

- Task coordination: *time planning* (TCT) are coordinative statements with reference to time,

time planning problem (TCTP) are statements signalling that a member or the group is not performing as planned, *task division* (TCD) statements aim to coordinate the tasks of group members, *task division problem* (TCDP) statements signal that a member or the group is not performing according to the task division, and *asking moderator* (TCM) are units in which a member or the group asks the moderator for advice;

- *Task content* (TCN) are units in which the content of the task is discussed;
- Non task statements are units that focus on social or technical issues: *positive social* (NTS+) are statements expressing a positive attitude toward a member or the group, *negative social* (NTS-) statements express a negative attitude toward a member or the group, and *technical* (NTT) statements are units in which technical issues or problems are expressed;
- *Non codable* (NOC) are units that cannot be assigned any other code.

The reliability of segmentation was computed as the proportion agreement because there is only one category involved with two values (agree = 1, disagree = 0); two or more categories requires computation of Cohen's kappa to correct for chance agreement. In addition, calculating kappa for segmentation reliability can easily lead to negative values (see Weinberger & Fischer, this issue). A threshold for the proportion agreement reliability of segmentation does not exist in CSCL research nor in the domain of content analysis (see Neuendorf, 2002; Riffe, Lacy, & Fico, 1998; Rourke et al., 2001). Thus, the proportion agreement threshold for coding used in content analysis is the most applicable where "a minimum level of 80% is usually the standard" (Riffe et al., 1998, p. 128).

Reliability of segmentation and coding

Two coders (A and B) received half an hour of training in the segmentation procedure, followed by segmenting a selection of 20 e-mail messages (12.5 % of a total of 160 messages),

resulting in a proportion agreement of 41% – well below the 80% threshold. Next, a set of 20 e-mails segmented by one of the principal investigators was coded by two coders (C and D) who received two hours of training with the coding categories. Cohen's kappa proved to be .30 (.60 is regarded as a minimum requirement; cf. Landis & Koch, 1977). It proved too difficult to separate segmentation and coding. It was decided to combine segmentation and coding (this time using communication sampled from the 'policy development' course). Still, the reliability of combined segmentation and coding remained unsatisfactory (Cohen's kappa .45).

Conclusion

It was assumed that the coding categories were not sufficiently developed or precisely defined and therefore too difficult to distinguish. For example, how should statements be coded in which 'time planning' and 'task division' occur simultaneous? What are indicators for a 'problem' that students experience? However, although the coding scheme could be improved, reflection on a possible explanation for the disappointing reliability of the procedure revealed a methodological problem with respect to the chosen unit of analysis (unit of meaning): unit boundary overlap.

Although studies that used a 'unit of meaning' for all coding (Veldhuis-Diermanse, 2002) or as part of a variable set of units (e.g., Schellens & Valcke (2002) used both messages and a 'unit of meaning' whenever applicable) report high intercoder reliabilities, it is likely that – due to the combined segmentation and coding – variations in the length of the 'unit of meaning' result in overlapping units with different codes. If these codes are treated as 'mutually exclusive' – as is the case in the 'quantitative content analysis' approach – this causes a serious methodological problem which is illustrated in Figure 1.

Insert Figure 1 about here

Figure 1 depicts the independent segmentation of two coders. Horizontal lines represent sentences in the communication transcript, brackets signal the start and end of a segment and the numbers represent a coding category. Coder A assigns the grey area to coding category 1, coder B thinks it belongs to coding category 2. Is the communication in the grey area trivial or not? Is the grey area a sentence, several sentences, or a paragraph? To date, proponents of the 'unit of meaning' have failed to address these questions.

To overcome the methodological problem of overlapping unit boundaries, the relative amount of 'communication' in the transcript that was assigned to a different segment and also received a different code (grey area in Figure 1) should be determined. Another solution would be to use a smaller unit to minimise the relative amount of the total communication that received a different – mutually exclusive – code by two independent coders. However, unit boundary overlap is not the only factor involved: the applicability of a unit is also affected by four contextual constraints

Four contextual constraints

Apart from the 'unit boundary' problem the applicability of a unit that is smaller than a message is also affected by four constraints: a) the object of the study, b) the nature of communication, c) the collaboration setting, and d) the technological communication tool used.

With respect to the *object of the study*, the difference between manifest and latent variables is most important (Neuendorf, 2002). Whereas 'quantitative' content analysis attempts to answer either descriptive or experimental research questions derived from manifest variables (similar to a prospective view), the object of 'qualitative' content analysis are latent variables that cannot be directly observed and need to be inferred from a transcript (retrospective view), e.g., 'knowledge construction' expanding over segments and messages, which is more susceptible to subjectivity

and more difficult to replicate. In sum, prior to any type of content analysis it is important that the variables of interest are determined (manifest or latent).

In general, the *nature of communication* is easily visible in the difference between transcribed verbal and written communication, as well as between written synchronous and asynchronous communication. Rourke et al. (2001) indicate that in asynchronous text-based messages (e.g., forum) 'telegraphic' and 'oral' styles were intermixed, and compound sentences are frequently used. In comparison, chat discussions resemble 'oral' communication in the sense that utterances are short and similar to speech utterances. Finally, the educational level can affects the communication style: in primary education most messages tend to be short and on a single topic (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003), whereas in higher education messages are far more complex and contain several topics that can belong to different coding categories.

The *collaboration setting* also affects the applicability of a unit. In a forum discussion on a specific topic, statements focus mainly on the topic of the discussion (Schellens & Valcke, 2002; Veldhuis-Diermanse, 2002). Here, comparatively little coordination is observed, whereas in a project-based collaboration setting, coordinative statements (e.g., task division, deadlines) will be frequently observed as the collaboration occurs in a prolonged period (months) compared to discussion of a topic (one or two weeks). In a project-based setting it can be more difficult to distinguish qualitatively different statements because these are mixed in a single unit.

Finally, the *technological tool* influences a unit's applicability. Synchronous chat evokes shorter statements than an e-mail or a threaded forum and the exchange rate of contributions is much higher in chats than in a forum. Also, design of the chat-tool design facilitates smaller communicative statements, whereas e-mail or threaded forums tend to force the participants to address multiple issues in a message. Finally, video conferencing places a specific demand as it

involves oral and non-verbal communication.

It can be concluded that choosing a unit of analysis can be guided by 'accepted practices', but the unit should be clearly defined and 'unit boundary overlap' should be limited and computed. A unit of analysis that is smaller than a message cannot be applied to the study of every research objective, nature of communication, collaboration setting and technological communication tool. These four contextual constraints should be taken into account prior to selecting or developing a content analysis procedure. Reflection regarding 'unit boundary' and the contextual constraints lead to defining an alternative unit of analysis, as well as a revised segmentation procedure and coding categories, for the analysis of the research context discussed.

Developing a content analysis procedure: alternative approach

Reviewing the coded transcripts used for the original procedure revealed that the communication had been subject to 'unit boundary overlap'. In addition, with respect to the four constraints the research objective focused on an experimental comparison involving 'manifest variables'. It was also apparent that e-mail communication combined 'oral' and 'telegraphic' styles: where a pause would occur in natural speech punctuation appeared and compound sentences were a rule rather than an exception. Messages were long and summations were observed. Given the project-based collaboration setting the students addressed multiple issues in a compound sentence. The original segmentation procedure provided no guidance to distinguish such small segments. Finally, some of the coordination categories ('time' and 'task division') posed coding problems.

It was concluded that the use of a 'unit of meaning' appeared to be too ambiguous to enable the coding of interest. Nevertheless, the objective to code and quantify the communication for statistical comparison remained. One approach to decrease 'unit boundary overlap' is using a smaller unit, such as a sentence or a proposition. Although smaller segments increase the number

of segments that cannot be coded, the accuracy of coding is improved because the ambiguity of the content is reduced. In other words, sentences or parts of compound sentences will more likely contain a single concept, expression or statement. Using a 'proposition' would be most adequate, however, such an analysis is very time consuming because it requires that all communication is re-written into single propositional utterances (Van Dijk & Kintsch, 1983).

Procedure

It was decided to develop an alternative segmentation procedure that would be systematic and independent of the coding categories. Although a sentence as unit of analysis is not uncommon (e.g. Fahy, Crawford, & Ally, 2001; Hillman, 1999), segmentation of compound sentences was added. The unit of analysis was thus defined as: a sentence or part of a compound sentence that can be regarded as 'meaningful in itself, regardless of the meaning of the coding categories'. Note that 'meaningful' is here used in a sense that is very different from its use in relation to a 'unit of meaning'. Punctuation and the word 'and' are 'markers' to segment compound sentences if the parts before and after the marker are a 'meaningful' sentence. This procedure is practical, not laborious and can be mastered in a single day. Table 2 depicts this procedure.

Insert Table 2 about here

Reliability of segmentation

An initial reliability test was performed on forty messages that were extracted from four groups (ten messages each). Two groups worked with predefined roles and two groups worked without them. The messages were independently segmented by two coders (A and B) who received only a written instruction (Table 2) and no explicit training. Each coder needed approximately ninety

minutes for segmentation. Afterwards, the proportion agreement (number of segments identified) was determined from the perspective of each coder, in contrast to other studies reporting a single proportion agreement (see also Prins et al., 1998). As shown in Figure 2 the selection of only one of either perspectives can significantly affect the proportion agreement (0 = disagree, 1 = agree): from perspective A there is 50% agreement (upper bound) and from perspective B there is only 33% agreement (lower bound).

Insert Figure 2 about here

The proportion agreement for the number of segments identified (the perspective of each coder serving as an upper and lower bound of the 'true' agreement) had a lower bound of 79.33 % (coder A) and an upper bound of 85.39 % (coder B).

A second reliability test was performed on fifty messages, extracted from four groups. Twenty of the fifty messages were similar (ten messages from two role groups) to test the rules 4a, 4d and 5e, added after the first reliability test (see Table 2). Fourteen and sixteen messages were sampled from two other groups working without the roles. All messages were independently segmented by two coders (A and C). Proportion agreement had a lower bound of 85.50 % (coder A) and an upper bound of 88.05 % (coder C).

Finally, cross-validation of the alternative procedure was performed on a dataset of English communication. Students collaborated in a networked learning community (hosted in WebCT©) in small project groups during seven weeks, during a course in a MEd programme on E-learning (see De Laat & Lally, 2003). Sixty-five messages were sampled from an asynchronous forum. Examination of the communication revealed a crucial difference: whereas in Dutch a dash is

rarely used for punctuation it is common in English. In addition, symbols such as '(…)' or '…' were frequently used to indicate pauses that occur in natural speech.

Sixty-five messages were sampled: 22 in the first week, 24 in the third and fourth week and 19 in the sixth and seventh week. Messages were independently segmented by two coders (A and C). Reliability of the segmentation had a lower bound of 87.47 % (coder A) and an upper bound of 91.44 % (coder C). The degree of proportion agreement shows that the alternative procedure can be applied to other asynchronous settings and that it is not language specific.

Reliability of coding

Quantifying the communication for comparative statistical analysis requires that two independent coders code the same segments. As long as the reliability of the segmentation is sufficiently high (a minimum of 80 %, see p. 10) – to decrease the probability that differences in unit boundaries result in overlapping but different codes – in principle, it does not matter whether segmentation by coder A or B is coded. It is also possible to choose the segmentation of the coder that discerns the smallest units, since the aim is to be precise (Prins et al., 1998).

Revising the coding categories

Simultaneous to the development of an alternative segmentation procedure, the coding categories were refined. Due to the change in the grain size of the unit the categories had to be refined, as the operationalisation of codes needs to be aligned with the measurement aim of the analysis (see De Wever, Valcke, Schellens, & Van Keer, this issue). Definitions of the categories were made more explicit and rules for coding were developed (for more detail see Strijbos, 2004). Eleven iterations were performed to refine the coding categories (the most significant changes will be discussed compared to the original coding categories).

First, the categories ‘time planning problem (TCTP)’ and ‘task division problem (TCVP)’ were removed. Interpretation of what constituted a ‘problem’ was problematic, and these codes were substituted for ‘specific’ and ‘unspecific’ references to ‘time’ and ‘activity’ (activity was formerly defined as task division), resulting in four subcategories. Two subcategories, specific and unspecific, were added for statements that combined ‘time’ and ‘activity’. Also, a category was added for ‘general coordination’. Furthermore, the ‘content’ category was expanded with two subcategories, one referring to statements about the ‘assignment’ and one to statements on ‘editing of their report’. Finally, a new main category was added for social statements about task performance with three subcategories that stress the focus of the statement: ‘general’, ‘towards an individual’ and ‘towards the group’. An overview of the final scheme consisting of five main categories and eighteen subcategories is shown in Figure 3. Mastery of the coding procedure is quite laborious. It takes about twenty hours of training: ten hours of coding and ten of discussion, in consecutive cycles of two hours with an experienced coder.

Cohen’s kappa was computed for three samples. Sample one consisted of 40 messages, samples two and three contained 50 messages. Each sample contained groups that worked with predefined roles and groups that worked without and was independently coded by two coders (A and B). Cohen’s kappa was computed for the subcategory level (sample 1, k = .62; sample 2, k = .60; sample 3, k = .63) and the main category level (sample 1, k = .69; sample 2, k = .70; sample 3, k = .68). Aggregation of the samples resulted in a kappa of .60 on subcategory level, which is considered satisfactory. For the main category level kappa was .70; considered to be substantial (cf. Landis & Koch, 1977).

Conclusion

The alternative unit of analysis can be determined reliably using the alternative procedure. This procedure is easy to use, not laborious and more precise. Cross-validation on an English dataset revealed that – after adding two 'punctuation' markers – the procedure was reliable for English datasets as well. Three samples were coded according to the revised categories, definitions and coding rules resulting in a satisfactory kappa for the subcategory level and a substantial kappa for the main category level.

Discussion

The initial aim of the research presented in this article was to develop a procedure for reliable content analysis of electronic communication in the context of project-based learning, however several issues emerged while developing this procedure that are not addressed in most CSCL research, but which have important implications for content analysis methodology and practice.

A review of CSCL 2001, 2002 and 2003 proceedings revealed that a considerable number of reports are vague in their definition of the unit of analysis used. Moreover, argumentation for choosing a specific unit of analysis, as well as a reliability statistic for coding – let alone for the segmentation if the unit is smaller than a message – is often not provided. Finally, the coding categories are briefly described and coding rules are not made explicit. Apart from a clear need for methodological rigour in content analysis, as well as reporting the reliability, it is apparent that the limitations of the applicability of a unit smaller than a message are barely addressed.

Practical experience with the 'unit of meaning' revealed a methodological issue that was defined as 'unit boundary overlap'. Supporters of the 'unit of meaning' argue that separating segmentation and coding is irrelevant because segmentation depends on the coding categories, however the 'unit boundary overlap' problem has clearly revealed that this argument is beside the point. Independent coders can assign different borders and codes, thus a reliability statistic

must be reported for both. Therefore, the questions regarding segmentation and coding in the case where the unit is smaller than a message (specifically the 'unit of meaning') still remain to be addressed (see Figure 1). In addition four contextual constraints were identified that affect the applicability of a unit of analysis smaller than a message. Practical experience with the 'unit of meaning' and re-examining the transcribed communication according to the 'object of research', 'nature of communication', 'collaboration setting' and 'technological tool', revealed a necessity to develop an alternative unit of analysis and segmentation procedure. It has been shown that this procedure – as well as its cross-validation – proved to be sufficiently reliable.

Thus, although a specific methodology or approach to content analysis may be regarded as an 'accepted practice' it is important that researchers are explicit about decisions taken during the development of a segmentation and/or coding procedure. The use of 'accepted practice' should not distract from criticism and rigour with respect to reliability. Researchers should beware that 'accepted practice' does not evolve into 'non-criticised practice'. With respect to 'quantitative' content analysis, intercoder reliability of both segmentation *and* coding are crucial to warrant the professed objectivity, reliability and replication. Especially if the research question requires the use of mutually exclusive categories to construct manifest variables for statistical comparison of experimental conditions. The procedure to determine the units and rules guiding coding should be systematically described, i.e. either it should be reported how much of the total number of segments and/or coding overlap or they should be performed as separate steps in the analysis, and a reliability statistic should be provided for both.

A final issue concerns the computation of reliability. Proportion agreement is appropriate with respect to the segmentation. Regarding the coding categories, Cohen's kappa is mostly used, but the issue of the number of categories included is ignored. The higher the number of categories,

the smaller the likelihood of chance agreement as opposed to possible deviations. In other words, kappa tends to be more strict (a comparatively larger chance agreement correction) in the case of fewer categories. For this reason we have reported two kappa statistics, one for the subcategory level and for the level of the main categories (for more detail on calculation of coding reliability see also De Wever, et al., this issue).

A scientific discourse is needed to answer the question 'What is an acceptable Cohen's kappa, given the research objective, nature of communication, collaboration setting and technical tool?'. Such a discourse could begin by introducing conventions for systematic reporting of coding and segmentation reliabilities and procedures. The procedures should be made available for cross-validation studies and secondary analysis (through reports or websites) to aid the establishment of construct validity (i.e., the operationalisation of the variable of interest, relative to the data) and the validity of the content analysis procedure with respect to different educational settings (Rourke & Anderson, 2004). Finally, if a previously validated coding scheme is used or adapted, the kappa of the original and adapted schemes should be reported.

In sum, based on the presented experiences in developing a content analysis procedure, five steps are recommended: 1) determine the unit of analysis given the probability of 'unit boundary overlap' and the four contextual constraints, 2) develop a segmentation procedure, 3) determine the reliability of the segmentation procedure, 4) re-use and/or adapt – or if necessary develop – coding categories and rules, and 5) determine the reliability of the coding categories. Although the segmentation procedure could possibly be applied to the analysis of verbal protocols of text and reading comprehension (Prins et al., 1998), transferability of the segmentation procedure to other research settings may be limited. However, researchers are invited to test the alternative

unit and segmentation procedure and/ or report whether the contextual constraints were useful to guide their choice for an applicable 'unit of analysis' given their research context.

CSCL may be a still emerging paradigm in educational research. However, methodological discourse needs to be part of any research paradigm – be it still evolving or not. Hopefully, this article can serve as a starting point for such a discourse.

References

References marked with an asterisk indicate studies included in the current review.

*Archer, W., Garrison, D. R., Anderson, T., & Rourke, L. (2001). A framework for analysing critical thinking in computer conferences. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 59-68). Maastricht: University of Maastricht.

*Arnseth, H. C., Ludvigsen, S., Wasson, B., & Mørch, A. (2001). Collaboration and problem solving in distributed collaborative learning. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 75-82). Maastricht: University of Maastricht.

*Armitt, G., Slack, F., Green, S., & Breer, M. (2002). The development of deep learning during a synchronous collaborative on-line course [Electronic Version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 151-159). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aviv, R. (1999, August). *Educational performance of ALN via content analysis*. Paper presented at the Sloan ALN Summer workshops (16-18 August), University of Illinois, Urbana, Illinois.

*Baker, M. J., Quignard, M., Lund, K., & Séjourné, A. (2003). Computer-supported collaborative learning in the space of debate. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 11-20). Dordrecht: Kluwer Academic Publishers.

Benbunan-Fich, R., & Hiltz, S. R. (1999). Impacts of asynchronous learning networks on individual and group problem solving: A field experiment. *Group Decision and Negotiation*, *8*, 409-426.

Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*, 271-315.

De Wever, B., Valcke, M., Schellens, T., & Van Keer, H. (this issue). Content analysis schemes to analyze transcripts of online asynchronous discussion groups. *Computers & Education*.

*De Laat, M. (2002). Network and content analysis in an online community discourse [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 625-626). Hillsdale, NJ: Lawrence Erlbaum Associates.

De Laat, M., & Lally, V. (2003). Complexity, theory and praxis: Researching collaborative learning and tutoring processes in a networked learning community. *Instructional Science*, *31*, 7-39.

Dillenbourg, P., Eurelings, A., & Hakkarainen, K. (Eds.). (2001). *European perspectives on computer-supported collaborative learning*. Maastricht: University of Maastricht.

*Erkens, G., Jaspers, J., Tabachnek-Schijf, H., & Prangsma, M. (2001). Computer-supported collaboration in argumentative writing. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 205-212). Maastricht: University of Maastricht.

Elliot, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, *38*, 215-229.

Fahy, P. J., Crawford, G., & Ally, M. (2001, July). Patterns of interaction in a computer conference transcript. *International Review of Research in Open and Distance Learning*, *2*(1). Retrieved May 20, 2004, from http://www.irrodl.org/content/v2.1/fahy.html

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, *12*, 213-232.

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.

*Häkkinen, P., Järvelä, S., & Byman, A. (2001). Sharing and making perspectives in web-based conferencing. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 285-292). Maastricht: University of Maastricht.

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

Harasim, L. (1993). Collaborating in cyberspace: Using computer conferences as a group learning environment. *Interactive Learning Environments*, *3*, 119-130.

Harasim, L., Hiltz, S. R., Teles, L., & Turoff, M. (1995). *Learning networks: A field guide to teaching and learning online*. Cambridge, MA: MIT Press.

Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Research*, *25*, 443-448.

Henri, F. (1992). Computer conferencing and content analysis. In A. Kaye (Ed.), *Collaborative learning through computer conferencing: The Najaden papers* (pp. 117-136). London: Spinger Verlag.

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, *28*, 31-45.

Hillman, D. C. E. (1999). A new method for analyzing patterns of interaction. *The American Journal of Distance Education*, *13*(2), 37-47.

*Hume, T. R., & Järvelä, S. (2001). Metacognitive processes in problem solving with CSCL in mathematics. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 301-307). Maastricht: University of Maastricht.

*Kirschner, P., Van Bruggen, J., & Duffy, T. (2003). Validating a representational notation for collaborative problem solving. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 163-172). Dordrecht: Kluwer Academic Publishers.

*Komis, V., Avouris, N., & Fidas, C. (2003). A study on heterogeneity during real-time collaborative problem solving. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 411-420). Dordrecht: Kluwer Academic Publishers.

Koschmann, T. (1996). Paradigm shifts and instructional technology: An introduction. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 1-23). Mahwah, NJ: Lawrence Erlbaum Associates.

*Lally, V. (2001). Analysing teaching and learning interactions in a networked environment: Issues and work in progress. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 397-405). Maastricht: University of Maastricht.

*Lally, V., & De Laat, M. (2002). Cracking the code: Learning to collaborate and collaborating to learn in a networked environment [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 160-168). Hillsdale, NJ: Lawrence Erlbaum Associates.

*Lally, V., & De Laat, M. (2003). A quartet in E. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 47-56). Dordrecht: Kluwer Academic Publishers.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

*Law, N., & Wong, E. (2003). Developmental trajectory in knowledge building: An investigation. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 57-66). Dordrecht: Kluwer Academic Publishers.

*Lenell, E., & Stahl, G. (2001). Evaluating affordance short-circuits by reviewers and authors participating in on-line journal reviews. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 406-413). Maastricht: University of Maastricht.

*Ligorio, B., Minnini, G., & Traum, D. (2001). Interlocution scenarios for problem solving in an educational mud environment. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 414-420). Maastricht: University of Maastricht.

*Ligorio, M. B., Talamo, A., Simons, R. J. (2002). Examining synchronous tutoring in a synchronous virtual world [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 615-616). Hillsdale, NJ: Lawrence Elrbaum Associates.

*Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2001). Analyzing patterns of participation and discourse in elementary students' online science discussions. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 421-428). Maastricht: University of Maastricht.

Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning & Instruction*, *13*, 487-509.

Louca, L., Druin, A., Hammer, D., & Dreher, D. (2003). Students' collaborative use of computer-based programming tools in science: A descriptive study. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 109-118). Dordrecht: Kluwer Academic Publishers.

Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology*, *91*, 1-20.

*Mäkitalo, K., Salo, P., Häkkinen, P., & Järvelä, S. (2001). Analyzing the mechanism of common ground in collaborative web-based interaction. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 445-453). Maastricht: University of Maastricht.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage.

Mudrack, P. E., & Farrell, G. M. (1995). An examination of functional role behaviour and its consequences for individuals in group settings. *Small Group Research*, *26*, 542-571.

*Mulder, I., Graner, M., Swaak, J., & Kessels, J. (2003). Stimulating questioning behaviour. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 421-430). Dordrecht: Kluwer Academic Publishers.

*Muukkonen, H., Lakkala, M., & Hakkarainen, K. (2001). Characteristics of university students' inquiry in individual and computer-supported collaborative study process. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 462-469). Maastricht: University of Maastricht.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage publications.

Newman, D. R., Webb, B., & Cochrane, C. (1995). *A content analysis method to measure critical thinking in face-to-face and computer supported group learning*. Retrieved May 20, 2004, from http://www.qub.ac.uk/mgt/papers/methods/contpap.html

*Pata, K., & Sarapuu, T. (2003). Framework for scaffolding the development of problem based representations by collaborative design. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 189-198). Dordrecht: Kluwer Academic Publishers.

Prins, F. J, Busato, V., Elshout, J., & Hamaker, C. (1998). Een nieuwe bijdrage tot de validatie van het (meta)cognitieve deel van de Inventaris Leerstijlen (ILS) [A new contribution to the validation of the (meta)cognitive part of the Inventory Learning Styles (ILS)]. *Pedagogische Studieën*, *75*, 73-93.

Rasku-Puttonen, H., Eteläpelto, A., Arvaja, M., & Häkkinen, P. (2003). Is succesful scaffolding an illusion? – Shifting patterns of responsibility and control in teacher-student interaction during a long-term learning project. *Instructional Science*, *31*, 377-393.

Riffe, D., Lacy, S., & Fico, F. G. (1998). *Analyzing media messages: Using quantitative content analysis in research*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research & Development*, *52*, 5-18.

Schellens, T., & Valcke, M. (2002). Asynchrone discussiegroepen: Een onderzoek naar de invloed op cognitieve kennisverwerving [Asynchronous discussion groups: Investigating the influence on cognitive knowledge gain]. *Pedagogische Studieën*, *79*, 451-468.

*Schwarz, B., Neuman, Y., Gil, J., & Ilya, M. (2001). Effects of argumentative activities on collective and individual arguments. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 545-552). Maastricht: University of Maastricht.

*Seitema-Hakkarainen, P., Lahti, H., Iivonen, M., & Hakkarainen, K. (2002). Computer support for participatory design – a pilot study [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 652-653). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stahl, G. (Ed.). (2002a). *Computer support for collaborative learning: Foundations for a CSCL community*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stahl, G. (2002b). Contributions to a theoretical framework for CSCL [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 62-71). Hillsdale, NJ: Lawrence Erlbaum Associates.

Strijbos, J. W. (2004). *The effect of roles on computer-supported collaborative learning*. Unpublished doctoral dissertation, Heerlen, The Netherlands, Open University of the Netherlands.

Strijbos, J. W., Kirschner, P. A., & Martens, R. L. (2004). What we know about CSCL: And what we do not (but need to) know about CSCL. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL: And implementing it in higher education* (pp. 245-259). Boston, MA: Kluwer Academic Publishers

Strijbos, J. W., Martens, R. L., & Jochems, W. M. G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, *42*, 403-424.

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

*Suthers, D., Girardeau, L., & Hundhausen, C. (2003). Deictic roles of external representations in face-to-face and online collaboration. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 173-182). Dordrecht: Kluwer Academic Publishers.

*Svensson, L. (2002). Interaction repertoire in a distance education community [Electronic version]. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 648-649). Hillsdale, NJ: Lawrence Erlbaum Associates.

*Tosonoglu-Blake, C., & Rapanotti, L. (2001). Mapping interactions in a computer conferencing environment. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 609-616). Maastricht: University of Maastricht.

*van Aalst, J., & Chan, C. (2001). Beyond "sitting next to each other": A design experiment on knowledge building in teacher education. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 20-28). Maastricht: University of Maastricht.

*Van Amelsfoort, M., & Andriessen, J. (2003). Comparing graphical and textual preparation tools for collaborative argumentation-based learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 5-9). Dordrecht: Kluwer Academic Publishers.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

*Van Drie, J., Van Boxtel, C., Erkens, G., & Kanselaar, G. (2003). Supporting historical reasoning in CSCL. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 93-112). Dordrecht: Kluwer Academic Publishers.

*Van Oostendorp, H., & Juvina, I. (2003). Role of icons and chat boxes in computer supported collaborative learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 275-279). Dordrecht: Kluwer Academic Publishers.

Veldhuis-Diermanse, E. A. (2002). *CSCLearning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education*. Unpublished doctoral dissertation, Wageningen, The Netherlands, Wageningen University.

Wang, M., Laffey, J., & Poole, M. J. (2001). The construction of shared knowledge in an Internet-based shared environment of expeditions (iExpeditions). *International Journal of Educational Technology*. Retrieved May 20, 2004, from http://www.ao.uiuc.edu/ijet/v2n2/v2n2feature.html

Wasson, B., Ludvigsen, S., & Hoppe, U. (Eds.). (2003). *Designing for change in networked learning environments*. Dordrecht: Kluwer Academic Publishers.

Weinberger, A., & Fischer, F. (this issue). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*.

Appendix A: Overview of the analysis units reported in CSCL 2001, 2002 and 2003 proceedings.

CSCL 2001

Study	*Unit of analysis*	*Segmentation reliability*
Archer, Garrison, Anderson, & Rourke	message	1)
Arnseth, Ludvigsen, Wasson, & Mørch	unclear	2)
Erkens, Japsers, Tabachneck-Schijf, & Prangsma	episode	not reported
Häkkinen, Järvelä, & Byman	message	1)
Hume & Järvelä	message	1)
Lally	unit of meaning	not reported
Lenell & Stahl	message	1)
Ligorio, Minnini, & Traum	unclear	2)
Lipponen, Rahikainen, Lallimo, & Hakkarainen	message	1)
Mäkitalo, Salo, Häkkinen, & Järvelä	message	1)
Muukkonen, Lakkala, & Hakkarainen	proposition	not reported
Schwarz, Neuman, Gil, & Ilya	argument	not reported
Tosonoglu-Blake & Rapanotti	unit of meaning	not reported
van Aalst & Chan	message	1)

CSCL 2002

Study (long papers)	*Unit of analysis*	*Segmentation reliability*
Armitt, Slack, Green, & Breer	utterance	not reported
Lally & De Laat	unit of meaning	not reported
Study (short papers)	*Unit of analysis*	*Segmentation reliability*
De Laat	unit of meaning	not reported
Ligorio, Talamo, & Simons	unclear	2)
Seitema-Hakkarainen, Lahti, Iivonen, & Hakkarainen	proposition	unclear
Svensson	unclear	2)

CSCL 2003

Study	*Unit of analysis*	*Segmentation reliability*
Baker, Quignard, Lund, & Séjourné	argument	not reported
Kirschner, Van Bruggen, & Duffy	utterance	4)
Komis, Avouris, & Fidas	unclear	2)
Lally & De Laat	unit of meaning	not reported
Law & Wong	discourse	3)
Mulder, Graner, Swaak, & Kessels	unclear	2)
Pata & Sarapuu	unclear	2)
Suthers, Girardeau, & Hundhausen	proposition	not reported
Van Amelsfoort & Andriessen	unclear	2)
Van Drie, Van Boxtel, Erkens, & Kanselaar	utterance	not reported
Van Oostendorp & Juvina	unclear	2)

1) A message or note is a fixed unit that (in general) can be determined objectively and reliable.
2) The unit of analysis is unclear, thus a reliability measure would not make sense.
3) Each group has only one discourse.
4) Utterances were determined by turn taking, a fixed unit that (in general) can be determined reliable.

Acknowledgements

The authors thank Mimi Crijns and Ger Arendsen of the department for management sciences at the Open University of the Netherlands for their support and assistance in gathering the data, as well as Vic Lally (University of Sheffield, United Kingdom) and Maarten De Laat (University of Southampton, United Kingdom and University of Utrecht, The Netherlands) for providing a data sample to conduct the cross-validation of the segmentation procedure.

Table 1

Overview of unit types reported in CSCL 2001, 2002 and 2003 proceedings (largest unit on top).

CSCL 2001		CSCL 2002		CSCL 2003	
Unit type	Frequency	Unit type	Frequency	Unit type	Frequency
				discourse*	1
message*	7	message*	1		
meaning	3	meaning	1	meaning	1
argument	1			argument	1
		utterance	1	utterance	2
proposition	1	proposition	1	proposition	1
unclear	2	unclear	2	unclear	5

* The discourse and message (e.g., e-mail or forum contribution) are non-ambiguous fixed units.

Table 2

Alternative unit of analysis and segmentation procedure.

1. Each message is first segmented in sentences by using a ‘full stop’, ‘question mark’ or ‘exclamation mark’ that the author of the message has written.

2. Each sentence that is followed by a ‘full stop’ constitutes a segment, regardless whether a ‘finite form’ or ‘verb’ is missing.

3. Each compound sentence is split in segments using punctuation signs and symbols or signs that are used for punctuation purposes:

    a. Comma
    b. Semicolon
    c. Colon
    d. Brackets
    e. The word ‘and’
    f. Dash
    g. (…) or …

    Segmentation is always subject to the criterion that each part of that compound sentence can be regarded as a ‘meaningful’ sentence in itself (regardless of the coding categories).

4. When determining whether a part of a compound sentence can be regarded as a ‘meaningful’ sentence in itself, the following rules apply:
    a. It is allowed to ignore the words that form the collocation;
    b. It is not allowed to add mentally a ‘finite form’ or ‘verbs’, if it has not been written.
    c. It is not allowed to leave out words that are written;
    d. It is allowed to mentally rearrange the order of ‘verbs’ and ‘finite form’ to create a ‘meaningful’ sentence;
    e. In case parts of a compound sentence share a conditional relationship, those parts are not regarded as separate segments;
    f. Statements between brackets are often in a telegraphic style, and thus they are difficult to rearrange in a ‘meaningful’ sentence. If either the ‘finite form’ or ‘verb’ is missing, the statement between parentheses will be regarded as a separate segment. The statement is not regarded as a separate segment if both are missing;
    g. Citations and hyperlinks that are included in the message are segmented according to the previous rules and examples below (see point five of this procedure that addresses the handling of summations (including summations of hyperlinks);
    h. If an abbreviation is used in the middle of a sentence, the sentence is not split after the ‘full stop’ at the end of that abbreviation;
    i. An introductory statement, two or three words, is not regarded as a separate segment (even if placed as such by the author) and is added to the next

sentence that it introduces;
   j. An introductory sentence is regarded as a separate segment.

5. Segmentation of summations: textual and lists (or bullets):
   a. If the majority of statements in a summation can be regarded as a 'meaningful' sentence in itself, each statement is treated as a separate segment;
   b. If the majority of statements in a summation can not be regarded as a 'meaningful' sentence in itself, all statements are treated as one segment;
   c. If half of the statements in a summation can be seen as a 'meaningful' sentence in itself, all statements are treated as a separate segment;
   d. In case the introductory sentence of a summation can be regarded as a 'meaningful' sentence in itself, this sentence is regarded as a separate segment. If not, this sentence is added to the first statement of the summation;
   e. If the main point in a summation is divided in sub points (e.g. 2.1, 2.2 etc.), than the above rules (see a, b, c) apply. An exception is the 'claw construction' in a summation: the main point and sub points comprise separate segments and the sentences in between can be regarded as a 'meaningful' sentence in itself. They are not directly part of the summation and thus behave as an appropriation in a summation.

Figure 1

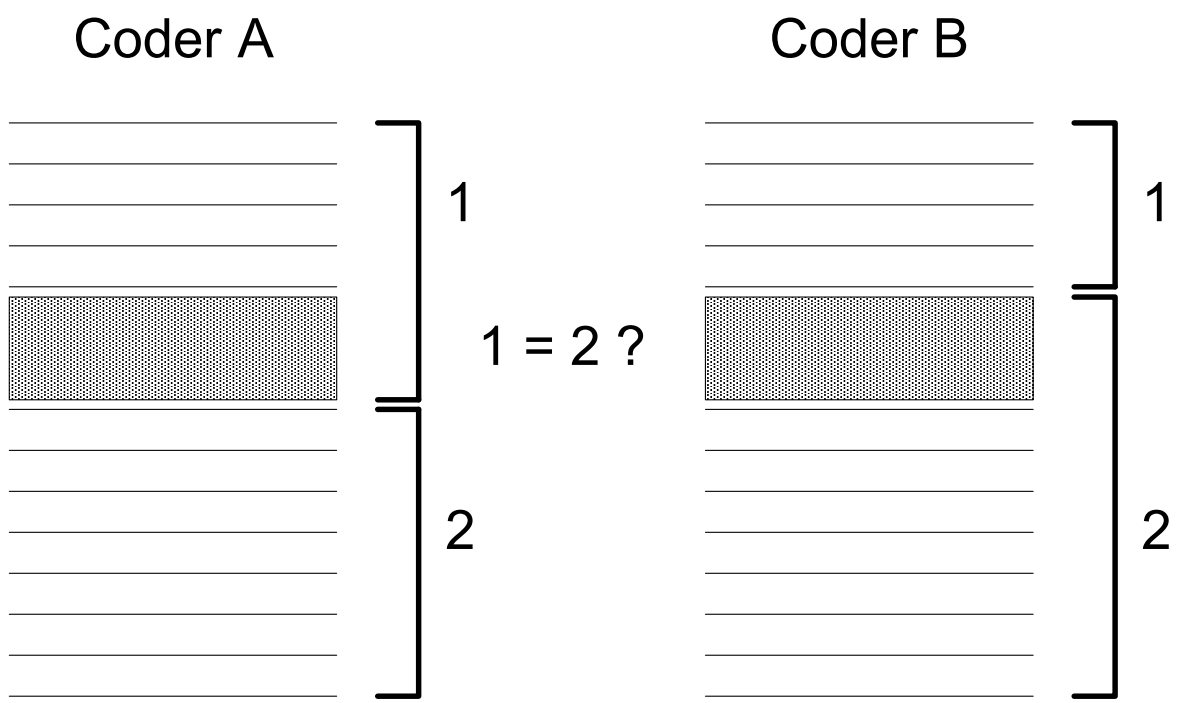

Figure 2

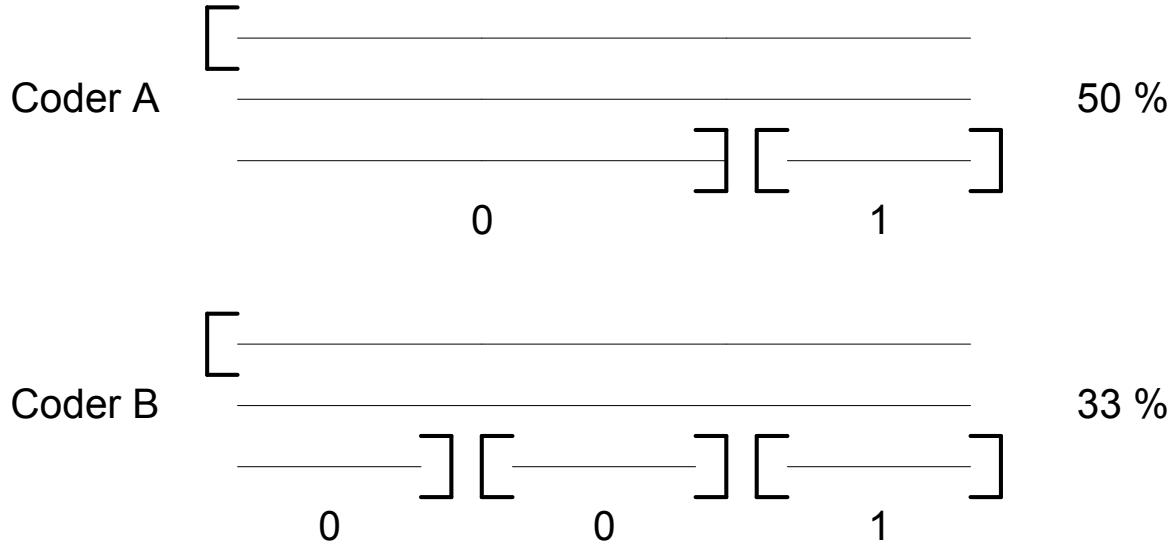

Figure 3

Code		Description	Example
Main	Sub		
TC	G	All statements with (a) a choice with no reference to time, the group or individuals; (b) coordination, but time nor activity, is indicated; (c) asking for a reaction but the object is unclear; (d) request a life sign from group members; (e) information on contextual factors that affects individual contribution to group work.	Why is nobody responding? Please give your ideas.
TC	TU	All types of statements regarding coordination in time, where time is indicated unspecifically.	I will be in touch again soon.
TC	TS	All types of statements regarding coordination in time, where time is indicated specifically.	I will be on holiday from June 8 until June 26.
TC	AU	All types of statements regarding coordination on activity, where the activity is (to be or was) performed by the group.	Who will make an inventory of all pressure groups involved?
TC	AS	All types of statements regarding coordination on activities or division of activities, where is indicated specifically who will perform that activity (i.e., by persons or by a (sub) group).	As far as I know, John Doe will perform the PERS analysis.
TC	TAU	All types of statements regarding coordination in time and activities or division of activities, where either time, division or both are indicated unspecifically.	I would like to know who will send me their comments on our report before Wednesday.
TC	TAS	All types of statements regarding coordination in time and activities or division of activities, where time and division are both are indicated specifically.	As agreed I expect that John Doe will send the analysis on Thursday.
TN	G	All types of statements that concern the general goal, or assessment criteria regarding the group assignment.	The assignment is about the public transport in Amsterdam.
TN	S	All types of statements that concern the content of the task (i.e., analysis of a policy problem) such as questions, comments, requests, providing information, information sources, content issues, discussion of that content, and so forth.	I believe that we have a different opinion about the interpretation of the analysis.
TN	R	All statements that concern the layout, structure and revision of the policy report.	We should delete section two and check for typing errors in three.
TS	G	All statements that concern general functioning or attitude toward the group, without reference to it or to individuals.	That's more like it!
TS	GR	All types of statements concerning group functioning, effort or attitude toward the group with reference to the group (i.e., use of we, all group members, or everybody).	I think we as group did a great job in a virtual project team.
TS	IN	All types of statements concerning an individual's functioning, effort or attitude toward another individual (i.e., with reference to names, he, she, I, you, they, (sub) Group 1).	John Doe, my compliments for your PERS analysis.
NT	A	All statements that concern the face-to face meeting at the start of the course and statements that concern acquaintance after the meeting (e.g., providing personal background information).	I have already met John Doe during the face-to-face meeting.
NT	T	All statements that concern technical issues (i.e., how to use, problems, evaluative remarks about computers, e-mail, specific software, and missing or forgotten attachments).	I am still struggling to find out how I am supposed to operate Edubox.
NT	S	All statements with a social orientation that are not related to the assignment (i.e., vacation, Christmas wishes).	How was your holiday in France?
NT	M	All statements wit an explicit reference to communication with the moderator or in which a group discusses the response.	We should ask the moderator if an analysis is useful.
NOC		All types of statements that not belong to any category specified (e.g., statements that signal receipt of a message or attachment).	Attached is a new schedule with the latest deadlines and tasks.

NOTE: Main = main code; Sub = subcode; TC = task coordination; TN= task content; NT= nontask; NOC= noncodable; G= general; TU= time unspecific; TS = time specific; AU = activity unspecific; AS = activity specific; TAU = time and/or activity unspecific; TAS = time and activity specific; S = social or specific; R = revision; GR = group; IN = individual; A = acquaintance; T = technical; M=

moderator.

Figure Captions

Figure 1. Unit boundary overlap.

Figure 2. Proportion agreement from two intercoder perspectives.

Figure 3. Abbreviated overview of the revised coding categories.