DOI:

10.1016/j.compedu.2005.04.005

http://dx.doi.org/10.1016/j.compedu.2005.04.005

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computers & Education xxx (2005) xxx–xxx

COMPUTERS & EDUCATION

www.elsevier.com/locate/compedu

Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review

B. De Wever *, T. Schellens, M. Valcke, H. Van Keer

Department of Education, Ghent University, H. Dunantlaan 2, 9000 Ghent, Belgium

Abstract

Research in the field of Computer Supported Collaborative Learning (CSCL) is based on a wide variety of methodologies. In this paper, we focus upon content analysis, which is a technique often used to analyze transcripts of asynchronous, computer mediated discussion groups in formal educational settings. Although this research technique is often used, standards are not yet established. The applied instruments reflect a wide variety of approaches and differ in their level of detail and the type of analysis categories used. Further differences are related to a diversity in their theoretical base, the amount of information about validity and reliability, and the choice for the unit of analysis.

This article presents an overview of different content analysis instruments, building on a sample of models commonly used in the CSCL-literature. The discussion of 15 instruments results in a number of critical conclusions. There are questions about the coherence between the theoretical base and the operational translation of the theory in the instruments. Instruments are hardly compared or contrasted with one another. As a consequence the empirical base of the validity of the instruments is limited. The analysis is rather critical when it comes to the issue of reliability. The authors put forward the need to improve the theoretical and empirical base of the existing instruments in order to promote the overall quality of CSCL-research.

Keywords: Computer-mediated-communication; Cooperative/collaborative learning; Distributed learning environments; Interactive learning environments; Learning communities

* Corresponding author. Tel.: +32 9 264 8676; fax: +32 9 264 8688.
E-mail address: bram.dewever@ugent.be (B. De Wever).

0360-1315/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2005.04.005

1. Introduction

Current educational practice reflects a growing adoption of computer tools to foster online collaboration. This practice is commonly described as the field of Computer Supported Collaborative Learning (CSCL). In CSCL-environments, online asynchronous discussion groups take a central place. These are known as Computer Mediated Conferencing (CMC), Computer Mediated Discussion (CMD), Computer Conferencing (CC), Networked Learning (NL), or Asynchronous Learning Networks (ALN). In spite of this conceptual variety, most environments have in common that students exchange messages through computers with one another. In this paper, we focus on text-based CSCL-tools. Next to email, discussion boards are the most commonly used tool in this context. Asynchronous text-based discussions present several advantages as compared to synchronous discussions: students get more opportunities to interact with each other and students have more time to reflect, think, and search for extra information before contributing to the discussion (De Wever, Schellens, & Valcke, 2004; Pena-Shaff & Nicholls, 2004). The fact that all communication elements are made explicit in the written contributions to the discussions "makes the process of collaboration more transparent [for the researcher], because a transcript of these conference messages can be used to judge both the group collaborative process and the contribution of the individual to that process [. . .]" (Macdonald, 2003, p. 378). All exchanges of information between students are stored in the discussion transcripts. These transcripts can be used by students for reflection purposes or they can serve as data for research (Meyer, 2004).

In the last decade, online asynchronous discussion groups have become a primary focus of educational research (Pena-Shaff & Nicholls, 2004). Researchers seem to agree that collaboration can foster learning (Lazonder, Wilhelm, & Ootes, 2003) and present a variety of theoretical frameworks to ground their assumptions (Schellens & Valcke, in press). Cognitive constructivists claim that the input in the CSCL-environment fosters learning due to the explicitation of individual knowledge elements (retrieval from memory) and the consecutive reorganization of knowledge elements in the course of the social transaction. Social constructivists argue that CSCL promotes the collaborative process in which meaning is negotiated and knowledge is co-constructed (Lazonder et al., 2003). Both views "acknowledge the importance of interaction in collaborative learning" (Lazonder et al., 2003, p. 292). This interaction, confined in the transcripts of the discussion, is thus the object of a large body of recent educational research.

At a first stage, research based on the discussion transcripts was mainly restricted to gathering quantitative data about levels of participation (Henri, 1992). However, these quantitative indices about numbers of student contributions hardly helped to judge the quality of the interaction (Meyer, 2004). At a later stage, content analysis was adopted as a technique to unlock the information captured in transcripts of asynchronous discussion groups. Therefore, Henri calls CMC a "gold mine of information concerning the psycho-social dynamics at work among students, the learning strategies adopted, and the acquisition of knowledge and skills" (1992, p. 118). Other researchers use the transcripts of online discussion to investigate the process of the social construction of knowledge (Gunawardena, Carabajal, & Lowe, 2001; Gunawardena, Lowe, & Anderson, 1997) or critical thinking (Bullen, 1997; Newman, Webb, & Cochrane, 1995). In general, the aim of content analysis is to reveal information that is not situated at the surface of the transcripts. To be able to provide convincing evidence about the learning and the knowledge construction that is taking place, in-depth understanding of the online discussions is needed.

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

The present study focuses on transcript analysis. This content analysis technique can be defined as "a research methodology that builds on procedures to make valid inferences from text" (Anderson, Rourke, Garrison, & Archer, 2001). Although this research technique is often used, standards are not yet established. The applied instruments reflect a wide variety of approaches and differ in their level of detail and the type of analysis categories used. Further differences are related to a diversity in their theoretical base, the amount of information about validity and reliability, and the choice for the unit of analysis. In the present article 15 content analysis instruments are discussed and research studies in which they have been applied are analyzed.

In order to present an overview of the current state-of-the-art, a number of instruments were selected, based on the following criteria: instruments applied, cited, or reflected upon in ISI-journals and CSCL-conferences, since these are the most important fora where scientific discussions about the development, use, and study of such instruments take place. Further, this selection was extended with recently developed instruments and instruments with a unique approach or a noticeable theoretical background. The list of instruments is not exhaustive, but reflects a balanced sample of what is currently used in the research field: Henri's model (1992); the model of Newman et al. (1995); the model of Gunawardena et al. (1997); the instrument of Zhu (1996); the instrument of Bullen (1997); the TAT of Fahy and colleagues (Fahy et al., 2000; Fahy, Crawford, & Ally, 2001); the instrument developed by Veerman and Veldhuis-Diermanse (2001); instruments for measuring cognitive, social, and teaching presence (Anderson et al., 2001; Garrison, Anderson, & Archer, 2000, 2001; Rourke, Anderson, Garrison, & Archer, 1999); the instrument of Järvelä and Häkkinen (2002); the instrument of Veldhuis-Diermanse (2002); the instrument of Lockhorst, Admiraal, Pilot, and Veen (2003); the instrument developed by Pena-Shaff and Nicholls (2004); and the instrument of Weinberger and Fischer (this issue).

Within the context of the present article, we discuss the quality of the analysis instruments, more specifically the theoretical background, the choice for a unit of analysis, and the reliability of the instruments. When available, we refer to other studies that applied the same analysis instrument. This helps to qualify the current state-of-the-art of CSCL-research based on content analysis instruments.

2. The quality of analysis instruments

Content analysis instruments should be accurate, precise, objective, reliable, replicable, and valid (Neuendorf, 2002; Rourke, Anderson, Garrison, & Archer, 2001). These criteria are strongly interrelated. *Accuracy* is the extent to which a measuring procedure is free of bias (nonrandom error), while *precision* is the fineness of distinction made between categories or levels of a measure (Neuendorf, 2002). Accuracy should be as high as possible, while precision should be high, but not exaggerated. *Objectivity* should be attained at all time (Rourke et al., 2001). Although interpretation is necessary and subjectivity might be unavoidable, one should be aware that subjectivity affects the *reliability* and the *validity* of studies. The latter is clearly related to the theoretical base of the studies and is discussed together with *replicability* in the next section. In subsequent sections, we elaborate further on the unit of analysis and the inter-rater reliability.

2.1. Theoretical base of the instruments

Although researchers seem to agree that collaboration can foster the learning process (Lazonder et al., 2003), there is no unambiguous theory available to guide research on computer mediated interaction (Stahl, 2003). Without a theoretical model of the collaborative learning process it is impossible to identify empirical indicators that will form the basis of a coding instrument as a standard against which to evaluate whether or not effective learning is occurring in the online discussions (Gunawardena et al., 2001). As Perraton (1988) argues: without a theoretical basis, research is unlikely to go beyond data gathering. The theoretical base is also of importance to ground the validity of the instruments. *Internal validity* focuses on the match between the conceptual definition and the operationalization (Neuendorf, 2002). This refers to *systematic coherence* which defines the relation between the theory and the models used. *External validity* is the possibility to generalize the findings to different settings (often called *generalizability*). This external validity can be supported by replications of (parts of) the research. Therefore, it is important to achieve high replicability (Neuendorf, 2002).

2.2. Unit of analysis

One of the issues under discussion is the choice of the unit of analysis to perform content analysis. Researchers can consider each individual sentence as a single unit of analysis (Fahy et al., 2001). A second option is to identify a consistent "theme" or "idea" (unit of meaning) in a message and to approach this as the unit of analysis (Henri, 1992). A third option is to take the complete message a student posts at a certain moment in the discussion as the unit of analysis (Gunawardena et al., 1997; Rourke et al., 2001). Every researcher has his or her reasons to choose for one of these possibilities, and there is not really an agreement. The choice for a unit of analysis is dependent on the context and should be well-considered, because changes to the size of this unit will affect coding decisions and comparability of outcome between different models (Cook & Ralston, 2003). In this respect, Schrire (this issue) refers to a dynamic approach in which data is coded more than once and the grain size of the unit of analysis is set, depending on the purpose and the research question. We refer to Strijbos, Martens, Prins, and Jochems (this issue) for a more in-depth discussion of the issue of unitization.

2.3. Inter-rater reliability

According to Rourke and colleagues (Rourke et al., 2001, p. 7) "the reliability of a coding scheme can be viewed as a continuum, beginning with coder stability (intra-rater reliability; one coder agreeing with herself over time), to inter-rater reliability (two or more coders agreeing with each other), and ultimately to *replicability* (the ability of multiple and distinct groups of researchers to apply a coding scheme reliably)." Inter-rater reliability is a critical concern in relation to content analysis. It is regarded as the primary test of objectivity in content studies and defined as "the extent to which different coders, each coding the same content, come to the same coding decisions" (Rourke et al., 2001, p. 6). Unfortunately, a large subset of studies do not report inter-rater reliability, which – according to Lombard, Snyder-Duch, and Bracken (2002) – "can be seen as the consequence of a lack of detailed and practical guidelines and tools available to researchers regarding reliability''. Next to reporting inter-rater reliability, it is also vital to report information about the training of the coders and the coding process. A clear and transparent coding procedure can guarantee the quality and the reliability of the research. In the next paragraphs, we elaborate on the calculation of the inter-rater reliability because it is a conditio sine qua non for content analysis.

There are a number of indexes used to report inter-rater reliability: percent agreement, Holsti's method, Scott's pi, Cohen's kappa, Krippendorff's alpha, Spearman rho, Pearson correlation coefficient, Lin's concordance correlation coefficient, Kupper-Hafner index, etc. (Krippendorff, 1980; Kupper & Hafner, 1989; Lombard et al., 2002; Neuendorf, 2002; Rourke et al., 2001). There is no general consensus on what index should be used. Below we discuss two coefficients that provide a good estimation on the inter-rater reliability.

Percent agreement is the result of the ratio between the number of codes which is agreed upon and the total number (agree + disagree) of codes. It is by far the most simple and most popular reliability index. It can accommodate any number of coders, but it has a major weakness: it fails to account for agreement by chance (Lombard et al., 2002; Neuendorf, 2002). Furthermore, the matching of the codes has to be very precise, codes that are close but not exactly the same result in disagreement. Holsti's method is a variation on this percent agreement index. However, it takes situations into account in which the two coders evaluate different units. When it is calculated across a set of variables, it is not considered as a good measure because it can veil variables with unacceptably low levels of reliability (Lombard et al., 2002).

Krippendorff's alpha is one of the three coefficients that account for chance agreement. The other two are Scott's pi and Cohen's kappa. Krippendorff's alpha is to be favored for several reasons. First, to calculate Scott's pi and Cohen's kappa, the only information taken into account is the nominal level of the data. Krippendorff's alpha takes into account the magnitude of the misses, adjusting for whether the variable is measured as nominal, ordinal, interval, or ratio (Krippendorff, 1980; Lombard et al., 2002; Neuendorf, 2002). Furthermore, it allows for any number of coders, whereas pi and kappa are only applicable for research based on two coders. Following Lombard and colleagues (Lombard et al., 2002), the ''biggest drawback to its use has been its complexity and the resulting difficulty of 'by hand' calculations, especially for interval and ratio level variables''. We do not consider this calculation as a major problem, since software exists to calculate this coefficient from the reliability data matrix (a matrix with for each coder the code he or she has given to the unit), for example R. R is available as freeware (http://www.r-project.org/).

As written above, there is no general agreement on what indexes should be used. Percent agreement is considered an overly liberal index by some researchers, and the indices which do account for chance agreement, such as Krippendorff's alpha, are considered overly conservative and often too restrictive (Lombard et al., 2002; Rourke et al., 2001). Therefore, we suggest calculating and reporting both indices. In this way, more information is given to the reader of research studies in order to judge the reliability. Interpretation of levels of inter-rater reliability is not straightforward, since there are no established standards available. There seems to be no real consensus for the percent agreement statistic. Often a cut-off figure of 0.75–0.80 is used; others state that a value of 0.70 can be considered as reliable (Neuendorf, 2002; Rourke et al., 2001). Also for chance correcting measures, no standard is available to judge the level of inter-rater reliability. When Cohen's kappa is used, the following criteria have been proposed: values above 0.75

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

(sometimes 0.80 is used) indicate excellent agreement beyond chance; values below 0.40, poor agreement beyond chance; and values in between represent fair to good agreement beyond chance (Krippendorff, 1980; Neuendorf, 2002).

Irrespective of the coefficients used, Lombard and colleagues formulate a number of guidelines. They identify the minimum information that should be provided (Lombard et al., 2002, p. 602):

- the size of and the method used to create the reliability sample, along with a justification of that method;
- the relationship of the reliability sample to the full sample;
- the number of reliability coders and whether or not they include the researchers;
- the amount of coding conducted by each reliability and non-reliability coder;
- the index or indices selected to calculate reliability and a justification of these selections;
- the inter-coder reliability level for each variable, for each index selected;
- the approximate amount of training (in hours) required to reach the reliability levels reported;
- where and how the reader can obtain detailed information regarding the coding instrument, procedures and instructions (for example, from the authors).

Only when all this information is reported, readers can make conclusions about the reliability of the instrument used in the context of a study. We consider it of crucial importance that more information about reliability is reported. It will advance the quality of research in the field of content analysis.

3. Discussion of instruments for content analysis

Rourke and Anderson (2003) suggest that instead of developing new coding schemes, researchers should use schemes that have been developed and used in previous research. Applying existing instruments fosters replicability and the validity of the instrument (Stacey & Gerbic, 2003). Moreover, supporting the accumulating validity of an existing procedure has another advantage, namely the possibility to use and contribute to a growing catalogue of normative data (Rourke & Anderson, 2003). In the CSCL-literature, many researchers do create new instruments, or modify existing instruments. Below, we discuss 15 of these instruments in order of development and publication. For each instrument, we focus on the scientific criteria discussed above: the theoretical framework, the unit of analysis, and the inter-rater reliability data. An overview is presented in Table 1.

3.1. Henri (1992)

One of the instruments most often cited and used as a starting point in many CSCL-studies, is the model of Henri (1992). Her instrument to analyze the transcripts of discussions is based on a cognitivist approach to learning; although she also refers to particular concepts, such as learning in a cooperative mode and to the concept of collective knowledge (Henri, 1992). A central concept in view of the content analysis instrument is interactivity. The definition of interactivity is borrowed from Bretz (1983), who states that interactivity is a three-step process: (1) communication

Table 1
Overview of the content analysis schemes

Instrument	Theoretical background	Unit of analysis	Inter-rater reliability
Henri (1992)	Cognitive and metacognitive knowledge	Thematic unit	Not reported
Newman et al. (1995)	Critical thinking	Thematic unit	Not reported
Zhu (1996)	Theories of cognitive and constructive learning – knowledge construction	Message	Not reported
Gunawardena et al. (1997)	Social constructivism – knowledge construction	Message	Not reported
Bullen (1997)	Critical thinking	Message (several indicators per message possible)	Percent agreement
Fahy et al. (2000)	Social network theory – Interactional exchange patterns	Sentence	Percent agreement Cohen's kappa
Veerman and Veldhuis-Diermanse (2001)	Social constructivism – knowledge construction	Message	Percent agreement
Rourke et al. (1999)	Community of inquiry – social presence	Thematic unit	Holsti's coefficient
Garrison et al. (2001)	Community of inquiry – cognitive presence	Message	Holsti's coefficient Cohen's kappa
Anderson et al. (2001)	Community of inquiry – teaching presence	Message	Cohen's kappa
Järvelä and Häkkinen (2002)	Social constructivism – perspective taking	Message – Complete discussion	Percent agreement
Veldhuis-Diermanse (2002)	Social constructivism – knowledge construction	Thematic unit	Percent agreement Cohen's kappa
Lockhorst et al. (2003)	Social constructivism – learning strategies	Thematic unit	Cohen's kappa
Pena-Shaff and Nicholls (2004)	Social constructivism – knowledge construction	Sentence (sometimes paragraphs)	Code-recode and interrater procedures, but no reported coefficients
Weinberger and Fischer (this issue)	Social constructivism – argumentative knowledge construction	The authors apply both units of analysis on micro-level and on macro-level	Percent agreement Cohen's kappa

of information, (2) a first response to this information, and (3) a second answer relating to the first.

The whole analytical framework of Henri (1992) consists of five dimensions: a participative, social, interactive, cognitive, and metacognitive dimension. The participative dimension comprises two categories: (1) overall participation, which is the total number of messages and accesses to the discussion and (2) the active participation in the learning process, which is the number of statements directly related to learning made by learners and educators. As she believes that messages of unequal length can not serve as precise measures of active participation, she proposes to divide messages into statements corresponding to units of meaning (Henri, 1992).

The social dimension comprises all statements or part of statements not related to the formal content of the subject matter. This operationalization is derived from the model of Berger, Pezdek, and Banks (1987) that states that social presence is at work in any statement not related to the formal content of the subject matter.

The interactive dimension is first divided in two parts: interactive versus non-interactive (independent) statements. Secondly, the interactive statements can be further subdivided into explicit versus implicit interactions. Furthermore, two different types of interactive messages are distinguished: responses and commentaries. This leads to five categories, namely (1) direct (explicit) responses, (2) direct (explicit) commentaries, (3) indirect (implicit) responses, (4) indirect (implicit) commentaries, and (5) independent statements.

The cognitive dimension consists out of five categories: (1) elementary clarification: observing or studying a problem identifying its elements, and observing their linkages in order to come to a basic understanding, (2) in-depth clarification: analyzing and understanding a problem which sheds light on the values, beliefs, and assumptions which underlie the statement of the problem, (3) inference: induction and deduction, admitting or proposing an idea on the basis of its link with propositions already admitted as true, (4) judgment: making decisions, statements, appreciations, and criticisms, and (5) strategies: proposing coordinated actions for the application of a solution, or following through on a choice or a decision. Furthermore, surface processing is distinguished from in-depth processing, in order to evaluate the skills identified.

The metacognitive dimensions comprise metacognitive knowledge and metacognitive skills. Metacognitive knowledge is declarative knowledge concerning the person, the task, and the strategies. Metacognitive skills refer to "procedural knowledge relating to evaluation, planning, regulation, and self-awareness" (Henri, 1992, p. 131). Henri does notice however that although the messages can reveal useful information, it is impossible to reveal the totality of the metacognitive processes. This means that "even if no metacognitive activity was noticed, one could not conclude that the students are weak in this area" (Henri, 1992, p. 133).

As Lally (2001, p. 401) points out: "One of the major strengths of Henri's approach to content analysis using categories is that it focuses on the social activity and the interactivity of individuals in a group at the same time as giving a picture of the cognitive and metacognitive processes of those individuals. However, one of its major limitations is that it gives us no impression of the social co-construction of knowledge by the group of individuals as a group, in a discussion or a seminar." Henri (1992) does not provide information about the code-recode reliability or the inter-rater reliability of her instrument. She did not empirically test the instrument. Although the instrument has been criticized (Bullen, 1997; Gunawardena et al., 1997; Newman et al., 1995; Pena-Shaff, Martin, & Gay, 2001; Pena-Shaff & Nicholls, 2004), it can be considered as pioneering work and has been the base for subsequent research.

The instrument was for example used in a study of Hara, Bonk, and Angeli (2000), involving 20 master and doctoral students in a 12 weeks course. The coding of 271 messages reflected a percent agreement of 0.78 for the social dimension, 0.75 for the cognitive dimension and 0.71 for the meta-cognitive dimension. McKenzie and Murphy (2000) applied Henri's model as a basis for their study, based on 157 messages from 25 students, working during 11 weeks. Based on a random sample of one-third of the messages, they report a percent agreement of 0.76 for the interactive dimension, 0.44 for the cognitive dimension and 0.95 for the analytical model that distinguishes in-depth processing from surface processing. Reanalyzing the data after collapsing the five categories of the cognitive dimension into only three categories resulted in a percent agreement of 0.68.

3.2. Newman et al. (1995)

The theoretical concepts that support the instrument of Newman et al. (1995) are group learning, deep learning, and critical thinking. The authors argue that there is a clear link between critical thinking, social interaction, and deep learning. They developed a content analysis instrument based on Garrison's (1991) five stages of critical thinking and Henri's (1992) cognitive skills. They identify 10 categories: relevance, importance, novelty, outside knowledge, ambiguities, linking ideas, justification, critical assessment, practical utility, and width of understanding. For each category, a number of positive and negative indicators are formulated and most indicators are fairly obvious opposites (Newman et al., 1995). A critical thinking ratio is calculated using the totals for each positive or negative indicator, with a minimum of −1 (all uncritical thinking, all surface-level learning) and a maximum of +1 (all critical thinking, all deep-level learning) (Newman et al., 1995). The authors adopt themes as the unit of analysis. The units may be phrases, sentences, paragraphs or messages illustrating at least one of the indicators. They only mark and count the obvious examples, and ignore less clear indicators (Newman et al., 1995). Furthermore, they claim that some indicators rely on subject knowledge and should therefore be identified by an expert in the domain. This makes it more difficult to involve multiple evaluators and limits control for subjective scoring. Although the authors urge others to replicate their work, they do not report reliability data and hardly information is presented about the empirical validation of the instrument. Marra, Moore, and Klimczak (2004) argue that calculating inter-rater reliability is not possible given that the unit of analysis varies from phrases, to paragraphs, or the entire posting.

3.3. Zhu (1996)

The theoretical framework of Zhu's study is based on a combination of Vygotsky's theory and theories of cognitive and constructive learning (Zhu, 1996). The zone of proximal development and the importance of social negotiation are put forward, together with the notion of reflective thinking of Dewey (1933). The instrument is based on the theory of group interaction of Hatano and Inagaki (1991) and the theory of question analysis of Graesser and Person (1994). Building on these theories, Zhu divides social interaction into vertical interaction, when ''group members will concentrate on looking for the more capable member's desired answers rather than contribute to

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

and construct knowledge'' (Zhu, 1996, p. 824) and horizontal interaction when ''members' desires to express their ideas tend to be strong, because no authoritative correct answers are expected to come immediately''. In relation to the latter, two types of questions are distinguished: type I questions or information-seeking questions are posed when information is missing, while type II questions or discussing questions are used to provide some kind of information, to seek opinions or to start a dialogue (Zhu, 1996). Other categories are answers, information sharing, discussion, comment, reflection and scaffolding. The category answers comprises messages with specific information in order to answer type I questions, while information sharing comprises more general information. Discussion refers to messages that focus on elaborating and sharing ideas. Comments refer to any non-interrogative statements concerning readings, while reflective notes focus on evaluation, self-appraisal, relating or linking messages, and adjusting learning goals and objectives. Scaffolding notes provide guidance or suggestions. Zhu (1996) uses entire messages as the units of analysis. She does not report information about the reliability of the coding scheme.

3.4. *Gunawardena et al. (1997)*

The instrument of Gunawardena et al. (1997) is presented as a tool to examine the social construction of knowledge in computer conferencing. It is based on grounded theory and uses the phases of a discussion to determine the amount of knowledge constructed within a discussion. The authors refer to the models of Henri (1992) and the model of Newman et al. (1995). They indicate that these models served as a useful starting point for analyzing asynchronous discussions, but that they are ''not very specific on how to evaluate the process of knowledge construction that occurs through social negotiation in CMC'' (Gunawardena et al., 1997, p. 402). The theoretical framework for the instrument results from social constructivist principles, more definitely the processes of negotiating meaning and coming to an understanding by discussing and contributing knowledge, thus resulting in the shared construction of knowledge (Kanuka & Anderson, 1998).

In an initial version of the analysis instrument, two types of learning were distinguished. First, a basic type of learning through which participants ''were active in each other's learning processes only by providing additional examples of concepts which in essence were already understood. This type of learning is called 'learning by accretion,' or pooling of knowledge'' (Gunawardena et al., 1997, p. 413). Second, a type of learning: ''that which actually required participants to adjust their ways of thinking to accommodate new concepts or beliefs inconsistent with their pre-existing cognitive schema'' (Gunawardena et al., 1997, p. 413).

This distinction was evaluated as too artificial (ibid, p. 413). It is at this point that they presented a model based on 5 phases ''reflecting the complete process of negotiation which must occur when there are substantial areas of inconsistency or disagreement to be resolved'' (ibid, p. 413).

In contrast to Henri (1992) and Gunawardena et al. (1997); Newman et al. (1995) use the entire message as the unit of analysis. Furthermore, they argue that knowledge construction evolves through a series of phases. The first phase is sharing and comparing of information, which comprises observations, opinions, statements of agreement, examples, clarifications, and identifications of problems. This is followed by phase 2: the discovery and exploration of dissonance or inconsistency among ideas, concepts, or statements. The third phase is negotiation of meaning and/or co-construction of knowledge, which includes negotiation, identifications of areas of agreement, and proposing new co-constructions on topics where conflict exists. The fourth phase is characterized by testing and modification of proposed synthesis or co-construction. These co-constructed statements are tested against existing cognitive schema, experiences, and literature. The fifth and final phase refers to statements of agreement and application of newly-constructed meaning, and encompasses summarizing agreements, applications of new knowledge, and metacognitive statements revealing new knowledge construction (Gunawardena et al., 1997; Kanuka & Anderson, 1998; Lally, 2001).

Lally (2001, p. 402) affirms that "the analytical model of Gunawardena and her colleagues contains several important features in terms of understanding teaching and learning in networked collaborative learning environments: (a) it focuses on interaction as the vehicle for the co-construction of knowledge, (b) it focuses on the overall pattern of knowledge construction emerging from a conference, (c) it is most appropriate in social constructivist and collaborative (student-centered) learning contexts, (d) it is a relatively straightforward schema, and (e) it is adaptable to a range of teaching an learning contexts."

With respect to the reliability of the coding scheme, Gunawardena et al. (1997) mention that the messages were coded independently by two researchers, but they do not report inter-rater reliability coefficients. They note that, in case of discrepancies, a single code was determined after discussion between the two coders, but they do not mention how often discrepancies have arisen.

Schellens and Valcke (2005) for example, applied this content analysis scheme to study the discussions of 230 students, during a 12 week undergraduate course. The percent agreement when coding the 1428 messages by three independent coders was 0.69. The analysis scheme was also linked to the analysis scheme of Veerman and Veldhuis-Diermanse (2001). The results of this analysis are discussed below.

Marra et al. (2004) employed the instrument of Gunawardena et al. (1997) and report a Krippendorff's alpha of 0.59 for the initial codes and 0.93 for "codes postinter-rater reliability discussions" (p. 31). They furthermore compared this model with the model of Newman et al. (1995) and argue that the former provides "a more holistic view of discussion flow and knowledge construction", whereas the latter provides "focused and segmented coding on certain potential indicators of critical thinking" (Marra et al., 2004, p. 39).

3.5. *Bullen (1997)*

Bullen's instrument focuses on critical thinking. The theoretical framework is based on different conceptualizations of this concept (Dewey, 1933; Ennis, 1987; Garrison, 1991). It is described as a purposeful mental process, involving a variety of cognitive and metacognitive skills. Critical thinking is reflective, evaluative, and reasonable (Bullen, 1997).

Bullen's instrument consists of four different categories of critical thinking skills. The analysis focuses on finding evidence of the use of these skills (positive indicators), and also on finding evidence of uncritical thinking (negative indicators). A ratio of positive indicators to negative indicators was used to determine the level of critical thinking of students. For the first category, clarification, positive indicators are: (a) focusing on a question, (b) analyzing arguments, (c) asking and answering questions of clarification, and (d) defining terms and judging definitions; while negative indicators are (a) focusing on a question unrelated to the problem, (b) analyzing arguments inappropriately, (c) asking inappropriate or irrelevant questions, or (d) incorrectly answering questions of clarification and incorrectly defining terms and inappropriately judging definitions. The positive indicators for the second category assessing evidence are (a) judging the credibility of a source and (b) making and judging observations; while the negative indicators are judgments and observations based on inappropriate criteria. The third category, making and judging inferences, has a long list of criteria for making and judging deductions, inductions, and value judgments as positive indicators, while negative indicators are making and judging inferences that do not follow the listed criteria. Positive indicators of the final category, using appropriate strategies and tactics, are for example using models, metaphors, drawings, and symbols to simplify problems or talking through a confusing issue with another person. Negative indicators are the inappropriate use of strategies and tactics. For the complete list of indicators and criteria, we refer to Bullen (1997).

Empirical research was based on a 14 week bachelor degree course, involving 13 students and 1 instructor. 207 messages were analyzed. Bullen reports data on the reliability of his instrument. Three coders were involved, but there was only 17% agreement between the three judges (Bullen, 1997). The scoring of one of the judges differed extremely due tot ambiguity in the indicators. The percent agreement was 0.58 when the scoring of the two other judges were compared (Bullen, 1997).

3.6. Fahy et al. (2000)

The theoretical context of the study of Fahy et al. (2001) is based on the definition of interaction of Gunawardena et al. (1997): "the totality of interconnected and mutually-responsive messages" (Fahy et al., 2001, p. 2). Fahy (2001, 2002a, 2002b) and Fahy et al. (2000, 2001) use a sentence in a message as the unit of analysis. They argue that the unit of analysis must be something obvious and constant within transcripts and that sentences are used to convey ideas.

Fahy et al. (2001) promote a holistic approach to transcript analysis. They apply the concept of a social network: social networks contain and are sustained both by context, and by the social interaction opportunities they offer. They focus on two network concepts: the structural and interactional exchange patterns observed in transcripts. Structural features are represented by the size (number of members), the density (ratio of the actual numbers of links to the possible total), and intensity (responsiveness and attentiveness of members to each other) of the social network. Interactional features include the kinds of content exchanged in the interaction and the exchange flow or the directness of the resulting interaction (Fahy et al., 2001). The interactional features are analyzed with the Text Analysis Tool (TAT). The TAT is based on the instrument of Zhu (1996). It distinguishes five categories: vertical questioning, horizontal questioning, statements and supports, reflecting, and scaffolding. At a later stage (Fahy et al., 2001) the TAT was updated by adding one category "References/authorities" that includes references, quotations, and paraphrases on the one hand and citations or attributions on the other hand.

The authors (Fahy et al., 2001) report reliability data based on three studies and involving three independent coders: (1) a code-recode intra-rater reliability of 86% agreement, (2) an inter-rater reliability of 60–71% agreement and (3) Cohen's kappa inter-rater reliability coefficient of 0.45–0.65. The studies build on the work of small groups of student ($n = 13$), working during about

15 weeks in a graduate course setting. Not the number of units of analysis is reported (sentences) but the number of words: 53671 words.

3.7. Veerman and Veldhuis-Diermanse (2001)

Veerman and Veldhuis-Diermanse situate the use of CSCL within a constructivist framework: "From a constructivist perspective, collaborative learning can be viewed as one of the pedagogical methods that can stimulate students to negotiate such information and to discuss complex problems from different perspectives"; furthermore "collaboration with other students provokes activity, makes learning more realistic and stimulates motivation" (Veerman & Veldhuis-Diermanse, 2001, p. 625). They present an analysis procedure for two categories of messages: task-related and not task-related messages. The categories reflect their specific interest in messages that contain explicit expressions of knowledge construction. They subdivide the task-related messages into three categories: new ideas (content not mentioned before), explanations (refining or elaborating already stated information), and evaluation (critical view on earlier contributions). They applied the instrument in four different settings (synchronous and asynchronous) and compared the outcomes, but they do not report information about reliability. Messages are the units of analysis, except for a single study, where messages were divided into separate contributions, depending on the theme of the content (thematic unit).

The authors applied the scheme in four consecutive studies, involving 40, 20, 30, and 14 students and during 6–12 weeks in the context of an undergraduate course. Large numbers of messages were analyzed (2040, 1287, 952, and 1088), but no information about reliability indices was made available.

Schellens and Valcke (2005) applied the model of Veerman and Veldhuis-Diermanse (2001) in a CSCL-setting involving 230 students during a 12 week first year university course. 1428 messages were coded by three independent coders. Assessment of inter-rater reliability resulted in quite high percent agreement measures. The initial value of this statistic was 0.81. Percent agreement for independent recoding after negotiation between the coders was 0.87.

De Laat and Lally (2004) analyzed discussions of a workshop in a fully virtual master's program in e-learning. The data consisted of the transcripts of discussions of 7 professionals, during three periods of 10 days (160 messages). They calculated a Cohen's kappa of 0.86, based on a 10% sample.

The research of Schellens and Valcke (2005) is one of the studies that tried to study the validity of the instrument by Veerman and Veldhuis-Diermanse (2001) by simultaneously coding the discussions using the instrument of Gunawardena et al. (1997). In this way, the authors could relate the theoretical position of both models (see Fig. 1). Category 1, 2, and 3 in the instrument of Veerman and Veldhuis-Diermanse (2001) relates respectively to phase 1 in the instrument of Gunawardena et al. (1997) whereas category 4 and 5 relates respectively to category 2 and 3. Both models are parallel to one another for the first three levels of knowledge construction. However, the coding scheme of Gunawardena et al. (1997) does not differentiate between lower cognitive processes. On the other hand, this scheme goes beyond the scheme of Veerman and Veldhuis-Diermanse (2001) and discriminates more advanced levels of knowledge construction, such as testing and applying newly constructed mental models.

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

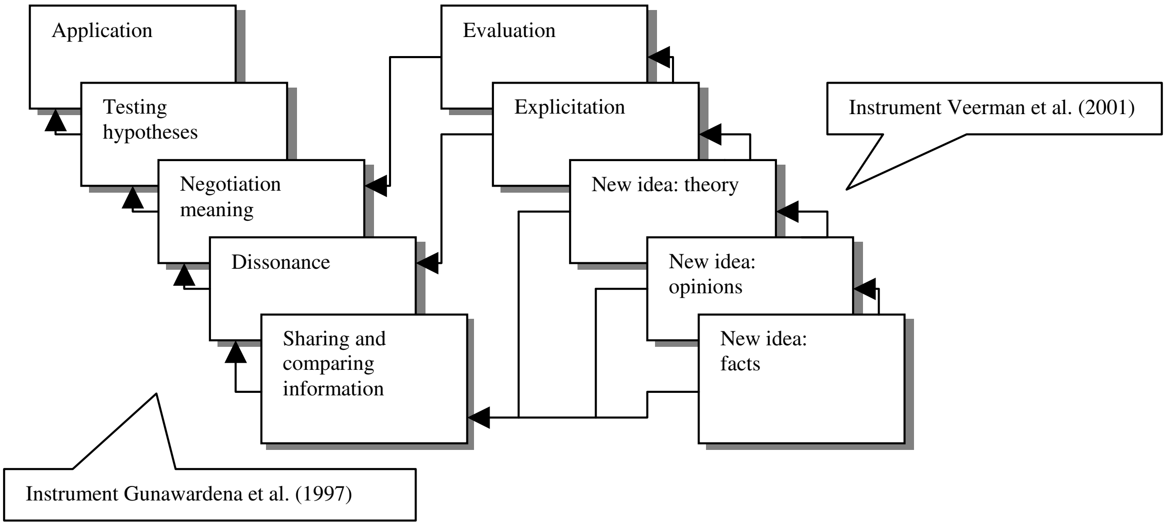

Fig. 1. Interrelations between two instruments to determine levels of knowledge construction (Schellens and Valcke, 2005).

3.8. Rourke et al. (1999)

Social presence is one of the three elements of the community of inquiry as conceptualized by Rourke et al. (1999). The other two elements are cognitive presence and teaching presence. It "supports cognitive objectives through its ability to instigate, sustain, and support critical thinking in a community of learners" (Rourke et al., 1999, p. 54). Social messages, such as jokes, compliments, and greetings do occur a lot in online asynchronous discussions (Rourke et al., 1999) and are considered to be important to motivate students. The social presence analysis model consists of three main categories: affective responses, interactive responses and cohesive responses. In their studies thematic units are used as the units of analysis. The authors claim that the units have the reliable identification attributes of syntactical units (Rourke et al., 1999). Two studies are reported in which the social presence analysis scheme was applied. Both studies were set up in the context of graduate level courses, involving 11–14 students, 2 moderator students and 1 instructor. A total of 90 and 44 messages were coded. The authors report Holsti's percent agreement indices from 0.91 to 0.95.

3.9. Garrison, Anderson, and Archer (2001)

Cognitive presence is another element in the community of inquiry model. "Cognitive presence reflects higher-order knowledge acquisition and application and is most associated with the literature and research related to critical thinking" (Garrison et al., 2001, p. 7). They operationalize cognitive presence through the practical inquiry process, which comprises four phases: (a) an initiation phase, which is considered a triggering event, (b) an exploration phase, characterized by brainstorming, questioning, and exchange of information, (c) an integration phase, characterized by constructing meaning and (d) a resolution phase, characterized by the resolution of the problem created by the triggering event (Garrison et al., 2001). Complete messages were chosen as the units of analysis. The model was tested in 2 empirical studies that lasted 13 and 2 weeks. A limited amount of students were involved: 11 students, 2 student moderators and 1 instructor. A total of 51 messages was analyzed. Varying levels of inter-rater reliability were reported: Holsti's coefficient of reliability (CR) of 0.45–0.84 and Cohen's kappa of 0.35–0.74.

3.10. Anderson et al. (2001)

Teaching presence is the third element in the overarching theoretical framework of the community of inquiry. The authors see "the function of the teacher as consisting of three major roles: first, as designer of the educational experience, including planning and administering instruction as well as evaluating and certifying competence; second, as facilitator and co-creator of a social environment conducive to active and successful learning; and finally, as a subject matter expert who knows a great deal more than most learners and is thus in a position to 'scaffold' learning experiences by providing direct instruction" (Anderson et al., 2001, p. 2). These three roles are the basis for their instrument to assess teaching presence. As unit of analysis the authors opt for the message, but they allowed "for the possibility that a single message might exhibit characteristics of more than one category" (Anderson et al., 2001, p. 11). Empirical research of the authors was based on a 13 week graduate level course, involving 1 instructor. A total amount of 139 students and 32 instructor messages were analyzed. Cohen's kappa inter-rater coefficients are reported and vary from 0.77 to 0.84.

3.11. Järvelä and Häkkinen (2002)

Järvelä and colleagues focus on three aspects: (a) the type of postings, (b) the level of discussions, and (c) the stage of perspective taking in discussions (Häkkinen, Järvelä, & Byman, 2001; Järvelä & Häkkinen, 2002). Their theoretical framework has its foundation in socio-constructivist learning theories in general, and more specifically in the idea of apprenticeship in thinking. With regard to the type of postings, the following categories are derived from the transcript data: (a) theory, (b) new point or question, (c) experience, (d) suggestion, and (e) comments. The message served as unit of analysis for this categorization. The concrete link between the analysis categories and the theoretical framework is not explained. No inter-rater reliability data when using this categorization are mentioned. Concerning the level of discussions, three categories are presented: (a) higher-level discussions, (b) progressive discussions, and (c) lower-level discussions. A complete discussion is considered as the unit of analysis for this categorization. An inter-rater agreement of 90% between two coders was reported. Negotiations resulted in a 100% consensus. The third aspect, stages of perspective taking in discussions, has been derived from Selman's (1980) perspective-taking categories. Selman (1980) defined five levels of the coordination of social perspectives, which served as a theoretical basis for the instrument, namely: (a) stage 0: undifferentiated and egocentric; (b) stage 1: differentiated and subjective role-taking; (c) stage 2: self-reflective, second person and reciprocal perspective; (d) stage 3: third-person and mutual perspective taking; and (d) stage 4: in-depth and societal-symbolic perspective taking. The unit of analysis for this aspect was again a complete discussion. Inter-rater agreement between two raters added up to 80%. Discussions between coders resulted in a 100% consensus (Järvelä & Häkkinen, 2002).

3.12. Veldhuis-Diermanse (2002)

Veldhuis-Diermanse (2002) developed a method to analyze students' learning in CSCL-environments. It is based on a constructivist view on learning and focuses on knowledge construction. More specifically it is partially rooted in the classification of Vermunt (1992), who distinguishes cognitive, affective, and metacognitive learning activities. Velduis-Diermanse's method consists of three steps. In a first step, the participation and interaction is analyzed. Both written and read notes are taken into account, together with the density of the discourse. The density is an indicator for the level of engagement in the discussions, and is measured by the proportion of actual connections between students to the maximum possible connections (Veldhuis-Diermanse, 2002). In the second step, the focus is on the different learning activities. This comprises cognitive learning activities, such as debating, using external or internal information; affective learning activities; and metacognitive learning activities, such as planning, keeping clarity, and monitoring. The third step focuses on the quality of constructed knowledge and is based on the structure of the observed learning outcome (SOLO) taxonomy of Biggs and Collis (1982), as described elsewhere in this issue (Schrire, this issue). Four levels are identified: level D (unistructural), where one relevant aspect of the task is picked up and used; level C (multistructural), where several relevant aspects of the task are acquired but not connected; level B (relational), where the learned components are integrated into a coherent whole; and finally the highest level A (extended abstract), where the acquired structure becomes transferable to the overall meaning (Veldhuis-Diermanse, 2002).

Meaningful units and whole messages were chosen as unit of analysis for respectively the first (step 2) and the second coding scheme (step 3). The author reports a Cohen's kappa of 0.82 (based on 20 randomly selected notes) for the analysis of cognitive learning activities (step 2) and a Cohen's kappa of 0.72 and percent agreement of 0.80 (based on 25 randomly selected notes) for the analysis of the quality of the knowledge constructed (step 3).

3.13. Lockhorst et al. (2003)

Lockhorst et al. (2003) base their instrument on a constructivist framework. They focus on online cooperation, and more specifically on the learning strategies that lead to an in-depth level of information exchange. They depart from the individual in the social state of affairs, and are less focused on the quality of the information exchanged or the knowledge constructed, but their main interest is the quality of the learning strategies used to construct knowledge.

The method developed by Lockhorst and colleagues is based on the analytical framework of Henri (1992). It includes five different instruments based on five perspectives. The first perspective is participation. This is measured by the number of statements and by Freeman's degree, which represents the centrality of a person in a social network. The second perspective is the nature of the content, which comprises four codes: (1) content related, (2) procedural, (3) social, and (4) no code. The third perspective is interaction and focuses on threads or chains of semantically or conceptually connected messages. For each thread the length, the number of layers, and the content is described. The fourth dimension focuses on information processing and is measured by a Likert scale from surface to deep information on a number of learning activities: (a) repeating, (b) interpreting, (c) argumentative, (d) adding new elements, (e) explaining, (f) judgmental, (g) asking questions, (h) offering solutions, (i) offering strategies and, (j) questioning. The fifth perspective is procedural information. Procedural statements are analyzed with an instrument that consists of six categories: (a) evaluative, (b) planning, (c) communication, (d) technical, (e) description, and (f) rest.

In accordance with Henri (1992), Lockhorst and colleagues use the unit of meaning as unit of analysis. For the second perspective (nature of content) a Cohen's kappa of 0.73 was calculated, comparing the work of two independent raters.

3.14. *Pena-Shaff and Nicholls (2004)*

Pena-Shaff and Nicholls (2004) developed an instrument to evaluate the knowledge construction processes in online discussions. Social constructivist learning theory served again as the theoretical framework for this instrument. The authors also concentrate on the quantitative analysis of participation and interaction rates (Pena-Shaff & Nicholls, 2004). Discussions with peers are considered to foster learning. The construction of knowledge is a social, dialogical process in which students should be actively involved (Pena-Shaff & Nicholls, 2004). Pena-Shaff and Nicholls (2004) make a distinction between 11 categories: question, reply, clarification, interpretation, conflict, assertion, consensus building, judgment, reflection, support and other. They further state that statements of clarification, interpretation, conflict, assertion, judgment, and reflection appear to be most directly related to the knowledge construction process.

The authors used sentences within messages as the basic unit of analysis, but also complete paragraphs are used as the unit of analysis, in order to maintain the meaning of a given sentence (Pena-Shaff & Nicholls, 2004). In their research, involving undergraduates, graduates, and university employees that worked together during 3 weeks, 152 messages of 35 students were analyzed. Coding and recoding was used to check for ambiguity in the coding. Two other independent coders were involved in the procedure. However, no reliability data have been reported.

3.15. *Weinberger and Fischer (this issue)*

Weinberger and Fischer (this issue) argue that learners in a CSCL-environment are often supposed to discuss their perspectives on a problem and engage in argumentative discourse with the goal to acquire knowledge. They propose a multi-dimensional approach to analyze argumentative knowledge construction. Four different process dimensions are identified: participation, epistemic, argumentative, and social mode. The participation dimension consists of two indicators, namely the quantity of participation, which designates whether learners participate at all, and the heterogeneity of participation, which specifies whether the learners participate on an equal basis. The epistemic dimension is divided into off-task and on-task discourse. The latter is further subdivided in three categories: the construction of problem space, the construction of conceptual space, and the construction of relations between conceptual and problem space. The argument dimension comprises the construction of single arguments, which encompasses claims, grounds with warrants, or qualifiers; and it comprises the construction of sequences of arguments, which includes arguments, counterarguments, and replies. The last dimension is the dimension of social modes of co-construction. It contains five categories: externalization, elicitation, quick consensus building, integration-oriented consensus building, and conflict-oriented consensus building. For an in-depth discussion of this framework, we refer to Weinberger and Fischer (this issue).

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

The authors apply units of analysis on both micro- and macro-level. A micro-segment contains a relation between two elements; these elements can be theoretical concepts or pieces of case information. Usually, micro-segments are a part of a sentence. A macro-segment consists of at least two micro-segments and is used to examine the relationship between these micro-segments. They report a percent agreement on micro-segmentation of 0.87, with a Cohen's kappa of 0.72. Furthermore, inter-rater reliability data for the different dimensions is available. A Cohen's kappa of 0.90 is reported for the epistemic dimension. For the argument dimension and the social modes dimension the authors report a Cohen's kappa of 0.78 and 0.81, respectively.

4. Discussion of the state-of-the-art in content analysis approaches

4.1. Theoretical framework

Stahl (2003) argues that the form of communication that appears in computer-mediated interaction "has special requirements and needs its own theory of communication". Studying the approaches discussed above, we can conclude that concepts from other theories or frameworks are borrowed, but a powerful theory to guide research is still lacking (De Laat & Lally, 2004; Stahl, 2004). When studying the theoretical frameworks of the instruments, a large variety of concepts are mentioned: cognitive and metacognitive knowledge and skills (Henri, 1992); critical thinking (Bullen, 1997; Newman et al., 1995); knowledge construction (Gunawardena et al., 1997; Pena-Shaff and Nicholls, 2004; Veerman and Veldhuis-Diermanse, 2001; Veldhuis-Diermanse, 2002; Weinberger and Fischer, this issue; Zhu, 1996); cognitive, social, and teaching presence (Anderson et al., 2001; Garrison et al., 2001; Rourke et al., 1999); perspective-taking (Järvelä & Häkkinen, 2002); interactional exchange patterns (Fahy et al., 2001); or learning strategies (Lockhorst et al., 2003). Although elements of the theoretical background are mentioned in all cases, not all studies present a clear link between the theory and the instruments. In this respect, the importance of systematic coherence is to be stressed. Some instruments elaborate the operational definition of theoretical concepts, while this is missing in other instruments. From the overview it is also clear that a number of researchers build on earlier work, but at the empirical level, links are hardly made between the new and previous analysis approaches.

A separate point of discussion is the differences between the instruments in the number of categories and the level of detail. Fahy et al. (2001) complain in this respect about the lack of discriminating capability of instruments. They are concerned that the communicative richness of transcripts may not be fully revealed when large portions of the transcripts are coded into very few interaction categories.

A last issue is the weak empirical base of the models. The majority of instruments has been developed in the context of limited empirical studies, building on small numbers of participants, restricted numbers of messages and discussions during short periods of time. Moreover, most empirical studies were descriptive in nature and did not primarily focus on hypotheses testing. This small research base does not favor the validation of the instruments nor does it help to underpin the theoretical foundation.

4.2. The unit of analysis

The unit of analysis determines how the overall discussion is to be broken down into manageable items for subsequent coding according to the analysis categories. The choice for the unit of analysis affects the accuracy of the coding and the extent to which the data reflect the true content of the original discourse (Hearnshaw, 2000). Four of the instruments discussed above use thematic units (units of meaning) (Henri, 1992; Lockhorst et al., 2003; Newman et al., 1995; Rourke et al., 1999). Seven recommend the use of complete messages as units of analysis (Anderson et al., 2001; Bullen, 1997; Garrison et al., 2001; Gunawardena et al., 1997; Järvelä & Häkkinen, 2002; Veerman & Veldhuis-Diermanse, 2001; Zhu, 1996). One study focuses on both thematic units and messages (Veldhuis-Diermanse, 2002) and another one uses micro- and macro-segments (Weinberger & Fischer, this issue). Only two studies use sentences as the unit of analysis (Fahy et al., 2001; Pena-Shaff & Nicholls, 2004). In one instrument, the whole discussion is the unit of analysis (Järvelä & Häkkinen, 2002).

The unit of analysis determines the granularity in looking at the transcripts in the online discussion. To get a complete and meaningful picture of the collaborative process, this granularity needs to be set appropriately. As is discussed in Strijbos, Martens, Prins, and Jochems (this issue) each choice represents advantages and disadvantages. It is striking that the choice for a specific unit of analysis is hardly linked to the theoretical base of the analysis instruments. What is for instance the best option when focusing on critical thinking? Most authors refer to criteria that are linked to objectivity and reliability in choosing the unit of analysis. The issue is however never related to validity questions. Garrison et al. (2000) indicate that opting for themes as the unit of analysis presents problems in terms of the reliable identification of each individual theme, resulting in subjectivity and inconsistency.

The fact that most studies opt for complete messages as the unit of analysis, is explained by the argument of Rourke et al. (2001) that this is the most objective identification of units of analysis, and that in this way researchers work with the unit as it has been defined by the author of the message.

Apart from the difficulties with regard to the choice of an appropriate unit of analysis, current reporting practices can be criticized. Most authors do not mention arguments for selecting or determining the unit of analysis; moreover a clear definition of the unit of analysis and the segmentation procedure is not always available and most of the studies do not report interrater reliability measures concerning the segmentation procedure (see also Strijbos et al., this issue).

4.3. Inter-rater reliability

The importance of a clear and transparent coding procedure and the inter/intra-rater reliability has been stressed throughout this article. We encouraged the use of multiple coefficients to determine inter-rater reliability, such as percent agreement and Krippendorff's alpha. Reporting multiple reliability indices is of importance considering the fact that no unambiguous standards are available to judge reliability values. Next to the concrete values, also information about the sample, the coding procedure, and the training should be reported carefully in order to improve the quality of research in the field of content analysis.

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

When studying the 15 instruments from this perspective, the picture is rather critical. In most studies, the procedure to determine the reliability is not reported. In five studies no reliability indices were reported. In two cases the authors reported Cohen's kappa inter-rater reliability coefficients, in four cases percent agreement or an equivalent measure was made available, and in four other cases both were reported. In order to give readers an overview of the inter-rater reliability of the coding schemes and procedures, calculating and reporting these measures is necessary.

5. Limitations and conclusions

The critical discussion of content analysis models, presented in this paper, has some limitations. Only a selection of content analysis instruments has been presented. Specific criteria were used to develop the list, but the overview is not complete. The same is true for the selection of studies that build on the work of the authors of the analysis instruments. Furthermore, we only discussed a basic set of criteria in relation to each instrument: the theoretical base, the unit of analysis, and reliability data. But these three aspects are crucial. The systematic coherence between theory and analysis categories, a grounded choice for the unit of analysis, and information about the (inter-rater) reliability and procedure are necessary conditions for applying content analysis in the context of a sound research methodology.

The picture that results from the analysis carried out in this paper is on some points unfavorable. As discussed above, coherent and empirically validated content analysis instruments are still lacking and so far these instruments have not fully resulted in progress in the development of the CSCL-research tradition. Therefore, the authors of the present article call for replication studies that focus on the validation of existing instruments in larger empirical studies. Hypothesis testing should be a central focus in these studies. The authors are convinced that this research reorientation will be helpful to foster the scientific quality and status of CSCL-research.

Acknowledgement

This study was funded in part by BOF2002/DRMAN/013.

References

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conference context. *Journal of Asynchronous Learning Networks, 5*http://www.sloan-c.org/publications/jaln/v5n2/pdf/v5n2_anderson.pdf (Retrieved: August 15, 2004).

Berger, D. E., Pezdek, K., & Banks, W. P. (1987). *Applications of cognitive psychology: Problem solving, education and computing*. London: Lawrence Erlbaum Associates.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic Press.

Bretz, R. (1983). *Media for interactive communication*. London: Sage.

Bullen, M. (1997). A case study of participation and critical thinking in a university-level course delivered by computer conferencing. Unpublished doctoral dissertation. University of British Columbia, Vancouver, Canada. Retrieved August 15, 2004, from http://www2.cstudies.ubc.ca/~bullen/Diss/thesis.doc.

Cook, D., & Ralston, J. (2003). Sharpening the focus: methodological issues in analysing online conferences. *Technology, Pedagogy and Education, 12*, 361–376.

De Laat, M., & Lally, V. (2004). It's not so easy: Researching the complexity of emergent participant roles and awareness in asynchronous networked learning discussions. *Journal of Computer Assisted Learning, 20*, 165–171.

De Wever, B., Schellens, T., & Valcke, M. (2004). Samenwerkend leren via informatie- en communicatietechnologie [Collaborative learning through ICT]. In I. D'haese & M. Valcke (Eds.), *Digitaal leren. ICT Toepassingen in het hoger onderwijs [Digital Learning. ICT applications in higher education]*. Tielt: Lannoo Campus.

Dewey, J. (1933). *How we think*. Boston, MA: D.C. Heath.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: theory and practice* (pp. 9–26). New York: W.H. Freeman.

Fahy, P. (2001). Addressing some common problems in transcript analysis. *International Review of Research in Open and Distance Learning, 1*(2). http://www.irrodl.org/content/v1.2/research.html/#Fahy (Retrieved August 15, 2004).

Fahy, P. (2002a). Assessing critical thinking processes in a computer conference. Centre for Distance Education, Athabasca University. Retrieved August 15, 2004, from http://cde.athabascau.ca/softeval/reports/mag4.pdf.

Fahy, P. J. (2002b). Coding transcripts: an illustration of issues and approaches. In J. Baggaley, P.J. Fahy, & C. O'Hagan (Eds.), *Educational conferencing: video and text traditions. Proceedings of the first international symposium on educational conferencing (ISEC), Bannff, Alberta*. Retrieved August 15, 2004, from http://cde.athabascau.ca/ISEC2002/papers/fahyMDDE.pdf.

Fahy, P., Ally, M., Crawford, G., Cookson, P. S., Keller, V., & Prosser, F. (2000). The development and testing of a tool for analysis of computer mediated conferencing transcripts. *Alberta Journal of Educational Research, 46*, 85–88.

Fahy, P., Crawford, G., & Ally, M. (2001). Patterns of interaction in a computer conference transcript. *International Review of Research in Open and Distance Learning, 2*. http://www.irrodl.org/content/v2.1/fahy.pdf (Retrieved August 15, 2004).

Garrison, D. R. (1991). Critical thinking and adult education: a conceptual model for developing critical thinking in adult learners. *International Journal of Lifelong Education, 10*, 287–303.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical thinking in a text-based environment: computer conferencing in higher education. *Internet and Higher Education, 11*, 1–14. http://communitiesofinquiry.com/documents/CTinTextEnvFinal.pdf (Retrieved August 15, 2004).

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education, 15*, 7–23. http://communitiesofinquiry.com/documents/CogPresPaper_June30_.pdf (Retrieved August 15, 2004).

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137.

Gunawardena, C. N., Carabajal, K., & Lowe, C. A. (2001). Critical analysis of models and methods used to evaluate online learning networks. In *American Educational Research Association Annual Meeting*. Seattle: American Educational Research Association.

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. Proceedings of the first european conference on CSCL*. Maastricht: McLuhan Institute, 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.

Hatano, C., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. Resnick, J. Levin, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331–349). Washington, DC: American Psychology Association.

http://dx.doi.org/10.1016/j.compedu.2005.04.005 - http://www.onderwijskunde.ugent.be/cscl/

Hearnshaw, D. (2000). Towards an objective approach to the evaluation of videoconferencing. *Innovations in Education and Training International, 37*, 210–217.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing. The Najadan Papers* (pp. 117–136). London: Springer-Verlag.

Järvelä, S., & Häkkinen, P. (2002). Web-based cases in teaching and learning: The quality of discussions and a stage of perspective taking in asynchronous communication. *Interactive Learning Environments, 10*, 1–22.

Kanuka, H., & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of Distance Education, 13*. http://cade.athabascau.ca/vol13.1/kanuka.html (Retrieved August 15, 2004).

Krippendorff, K. (1980). *Content analysis, an introduction to its methodology*. Thousand Oaks, CA: Sage Publications.

Kupper, L. L., & Hafner, K. B. (1989). On assessing interrater agreement for multiple attribute responses. *Biometrics, 45*, 957–967.

Lally, V. (2001). Analysing teaching and learning interactions in a networked collaborative learning environment: issues and work in progress. In *Euro CSCL 2001* (pp. 397–405). Maastricht McLuhan Institute. (Retrieved August 15, 2004), from http://www.mmi.unimaas.nl/euro-cscl/Papers/97.doc.

Lazonder, A. W., Wilhelm, P., & Ootes, S. A. W. (2003). Using sentence openers to foster student interaction in computer-mediated learning environments. *Computers & Education, 41*, 291–308.

Lockhorst, D., Admiraal, W., Pilot, A., & Veen, W. (2003). Analysis of electronic communication using 5 different perspectives. Paper presented at ORD 2003 (in Heerlen).

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.

Macdonald, J. (2003). Assessing online collaborative learning: process and product. *Computers & Education, 40*, 377–391.

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

McKenzie, W., & Murphy, D. (2000). I hope this goes somewhere: evaluation of an online discussion group. *Australian Journal of Educational Technology, 16*, 139–257.

Meyer, K. (2004). Evaluating online discussions: four different frames of analysis. *Journal of Asynchronous Learning Networks, 8*(2), 101–114.

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. *Interpersonal Computing and Technology, 3*, 56–77. http://www.qub.ac.uk/mgt/papers/methods/contpap.html (Retrieved August 15, 2004).

Pena-Shaff, J. B., Martin, W., & Gay, G. (2001). An epistemological framework for analyzing student interactions in computer-mediated communication environments. *Journal of Interactive Learning Research, 12*, 41–48. http://www.aace.org/dl/files/JILR/Jilr-12-01-41.pdf (Retrieved August 15, 2004).

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

Perraton, H. (1988). A theory for distance education. In D. Sewart, D. Keegan, & B. Holmberg (Eds.), *Distance education: International perspectives* (pp. 34–45). New York: Routledge.

Rourke, L., & Anderson, T. (2003). Validity in quantitative content analysis. (Retrieved August 1, 2004), from http://communitiesofinquiry.com/sub/papers.html.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education, 14*, 51–70. http://cade.athabascau.ca/vol14.2/rourke_et_al.html (Retrieved August 15, 2004).

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. (in press). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education*, doi:10.1016/j.compedu.2004.07.010.

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.

Schrire, S. (this issue). Knowledge-building in asynchronous discussion groups: going beyond quantitative analysis. *Computers & Education*, doi:10.1016/j.compedu.2005.04.006.

Selman, R. L. (1980). *The growth of interpersonal understanding*. New York: Academic Press.

Stacey, E., & Gerbic, P. (2003). Investigating the impact of computer conferencing: content analysis as a manageable research tool. In G. Crisp, D. Thiele, I. Scholten, S. Barker, & J. Baron (Eds.), *Interact, integrate, impact: Proceedings of the 20th annual conference of the australasian society for computers in learning in tertiary education, Adelaide, 7–10 December 2003*. (Retrieved September 1, 2004), from http://www.ascilite.org.au/conferences/adelaide03/docs/pdf/495.pdf.

Stahl, G. (2003). Communication and learning in online collaboration. In *Paper presented at GROUP'03, Sannibel Island, Florida*. (Retrieved June 26, 2003), from http://www.cis.drexel.edu/faculty/gerry/publications/conferences/2003/group/group03.doc.

Stahl, G. (2004). Building collaborative knowing: contributions to a social theory of CSCL. In J. W. Strijbos, P. Kirschner, & R. L. Martens (Eds.), *What we know about CSCL in higher education*. Amsterdam: Kluwer.

Strijbos, J. W., Martens, R., Prins, J., & Jochems, W. (this issue). Content Analysis: What are they talking about? *Computers & Education*, doi:10.1016/j.compedu.2005.04.002.

Veerman, A., & Veldhuis-Diermanse, E. (2001). Collaborative learning through computer-mediated communication in academic education. In *Euro CSCL 2001* (pp. 625–632). Maastricht: McLuhan institute, University of Maastricht.

Veldhuis-Diermanse, A. E. (2002). CSCLearning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education. Unpublished doctoral dissertation. Wageningen Universiteit, Nederland. (Retrieved August 15, 2004), from http://www.gcw.nl/dissertations/3187/dis3187.pdf.

Vermunt, J. D. (1992). Leerstijlen en sturen van leerprocessen in het hoger onderwijs. Naar procesgerichte instructie in zelfstandig denken [Learning styles and regulating of learning in higher education: Towards process-oriented instruction in autonomous thinking]. PH dissertation. Lisse: Swets & Zeitlinger.

Weinberger, A., & Fischer, F. (this issue). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, doi:10.1016/j.compedu.2005.04.003.

Zhu, E. (1996). Meaning negotiation, knowledge construction, and mentoring in a distance learning course. In: *Proceedings of selected research and development presentations at the 1996 national convention of the association for educational communications and technology*. Indeanapolis: Available from ERIC documents: ED 397 849.