

Methodological Issues in Analyses of Asynchronous, Text-Based Computer
Conferencing Transcripts

Liam Rourke

Terry Anderson

Randy Garrison

Walter Archer

University of Alberta

*Writing in any medium is an act of appropriation. Writing pulls words
or ideas out of their original time and stores them away for later use.
(Bolter, 1991, p. 42)*

Scenario

Prof. Jones has just completed first university course delivered entirely on-line. The 13 week semester class, has left Jones in a mild state of exhaustion. However, the course is finished, the marks have been assigned, and now thinks Jones, time for some reflection, analysis and perhaps a publishable paper. Jones smiles, confident in the knowledge that that the complete transcript of messages exchanged during the course has been captured in machine-readable format. She feels that this accessible data will confirm her hypothesis that students in the on-line course engaged in much higher level of discourse and discussion than any she had experienced in 10 years of face-to-face instruction. Further, she is interested in investigating the impact of the collaborative learning activity that she instituted in the middle of the course.

Jones is quickly disappointed. The 13 week discussion generated 950 messages. Merely reading them takes her four days. Attempts at cutting and pasting illustrations of higher level thinking into a word processor, have resulted in a hodge-podge of decontextualized quotations -each disparate enough to have Prof. Jones questioning her own definitions of higher order thinking. Realizing

that the analysis is going no where, Prof. Jones steps back to the literature and finds a set of criteria that define the broad areas of thinking skills she sees being developed in the transcripts. Heartened, but now running out of time Prof. Jones hires two graduate students to review the messages and identify the incidents of higher order thinking as defined by the expert. Two weeks later, the students report their results - not only have they failed to agree on 70% of the categorizations, but one student has identified 2032 incidents in the transcript, while the other found only 635 incidents. To add to her misery, Prof Jones also learns that her University ethics committee, concerned with the large increase in use of computer conferencing for credit courses, has ruled that without informed consent from students, her analysis, does not conform with guidelines of the university ethical research policy. Feeling overwhelmed and depressed, Prof. Jones returns to the education research once again, only to find that most of the issues she has been dealing with have not been addressed by major researchers in the field. She also finds that there is no coherent, long term tradition of researchers who have resolved the methodological problems inherent in text based, computer conference transcript analysis. This paper is written for the Prof. Jones's of the world hoping that it helps us to release the educational treasures that we believe are locked in the transcripts that document learning in the on-line environment.

Introduction

The unique capacity of computer conferencing to support interaction between and amongst participants while providing for temporal and spatial independence creates a uniquely valuable environment for distance, distributed, and lifelong learning applications. Additionally, the automatically-recorded and machine-readable data generated by this technology offers a unique opportunity for educational researchers. Many researchers describe their review of the transcript they are studying as 'content analysis' though few define exactly what they mean by the term, and fewer still pay heed to the long tradition of content analysis developed largely within the discipline of communications study.

Berelson (1952) defines content analysis as "...a research technique for the objective, systematic, quantitative description of the manifest content of communication" (p. 519). Many researchers characterize their review of conferencing transcripts as 'content analysis,' yet, their methodology does not always adhere to the criteria described by Berelson. Researchers using the technique have described it as difficult, frustrating, and time-consuming, and only a conspicuous few, notably Anderson (Kanuka & Anderson, 1998; Anderson & Kanuka, 1997; Gunawardena, Lowe, & Anderson, 1997; Anderson, 1996) and Newman (Newman, Webb, & Cochrane, 1995; Newman, 1994; Webb, Newman, & Cochrane, 1994) have repeated a content analysis effort.

This paper will explore the particular difficulties that researchers have encountered and the advances that they have made in the struggle to extract meaning from conferencing transcripts. It is not meant to be a meta-analysis of results, but rather a review of the methodology. Our intent is to document the evolution of content analysis techniques as they have been applied by us and other researchers to analyze transcripts of asynchronous, computer mediated conferencing in formal educational settings. Hopefully, this will facilitate the larger goal of improving the quality of teaching and learning using this medium. To accomplish this, we have reviewed a sample of 14 studies that are commonly referenced in the literature (see Table 1). We hope that this will provide subsequent researchers with a privileged starting point for their studies and refine the application of this powerful technique.

Characteristics of Content Analysis Methodology

Objectivity

Berelson's definition of content analysis begins with the term 'objective.' In the methodological context, we use this contentious term to refer only to the scientific practice of making verifiable observations of operationally defined constructs. The use of terms like "objectivity" seem somewhat anachronistic in the post modern era of educational research in the late 90's. We hasten to qualify

our understanding of "objective" to imply only that subsequent researchers, with adequate amounts of training will be capable of identifying the same units and categories of analysis as has been identified by earlier researchers. We do not believe that there is a single "objectivity" that resides outside of the interpretations of those who perceive it, but we do believe that human beings can and do learn to consistently describe and understand their external environment. Objectivity for us, is similar to the broader, more technical category of reliability.

Content analysis perhaps more so than any other quantitative technique, is susceptible to the infiltration of interpretive bias. Mower's (1996) frank reliability discussion is illustrative:

In ...instances of disagreement, [rater 1] agreed that [rater 2's] evaluation could be correct. In ...other instances of disagreement, it was determined that ... remarks could fit into either one of two categories depending upon the [rater's] interpretation. It was concluded that sometimes, subjective judgment was involved in assigning some topics to categories" (p. 220).

Mower's candor reveals a pervasive issue in content analysis studies. However, a quantitative study should not conclude with an admission that objectivity and reliability have not been achieved. Rather, the discovery of subjectivity should signal to the research team that further refinement is needed in category definition or coding protocol. Similarly, several studies state that inter-rater reliability was obtained through discussion between coders. These types of discussion represent an important stage in content analysis; however, reports of reliability should reflect the degree to which coders, working independently, have come to the same categorization.

Reliability

The first test of objectivity is reliability. Of the 14 published studies in our sample, only eight reported reliability (Fahy, Crawford, Ally, Cookson, Keller, & Prosser, 1999; Angeli, Bonk, & Hara, 1998; McDonald, 1998; Weiss & Morrison, 1998; Marttunen, 1997; Mowrer, 1996; Newman, Webb, & Cochrane, 1995; Ahern, Peck, & Laycock, 1992). Of these eight, two include exemplary discussions

of reliability (McDonald; Marttunen). The remaining six report interjudge agreement percentages. Statisticians characterize interjudge agreement as "...the most primitive approach to studying inter-rater agreement..." and "...clearly inadequate" because it does not account for chance agreement among raters (Capozzoli, McSweeney, & Sinha 1999; p. 5). Cohen's kappa, reported by McDonald and Marttunen is a chance-corrected measure that assumes two raters, n cases, and m mutually exclusive and exhaustive nominal categories. Meeting each of these assumptions drastically improves a study's validity. When confronted with a study that does not report acceptable interrater reliability figures, the reader must ask which of these critical assumptions was not met.

Replicability

In quantitative transcript analysis it is important that independent researchers be able to evaluate the same transcript and replicate the results of the analysis. This is not to suggest that we are replicating the creation of unique educational environments created during the course, but that we are replicating the analysis of the records of these educational encounters.

Reliability can be viewed on a continuum ranging from the meager coder stability to the robust replicability. Thus, replicability is the ultimate test of reliability. Reliability must be apparent within the coding and categorization of a single rater over time (intra-rater reliability) as well as between two independent raters (inter-rater reliability). It is important that independent researchers be able to evaluate the same transcript and replicate the results of the analysis. This is not to suggest the replication of the unique educational environments created during the course, but replication of the analysis of the records of these educational encounters.

The most commonly replicated protocol originates in Henri's (1991) seminal article. However, each time it is used, it is criticized and either modified or abandoned (Angeli, Bonk, & Hara, 1998; Bullen, 1998; Howell-Richardson & Mellar, 1996; Newman, Webb, & Cochrane, 1995; Kanuka & Anderson, 1998; Gunawardena, Anderson, & Lowe, 1997). The criticism that Henri's procedure

has received is, paradoxically, complimentary: Most researchers explicitly build upon the ideas enunciated by Henri, and no other model has generated significant replication. This lack of replication should be regarded as a serious problem. Even when a valid study yields statistically significant findings, the results are described cautiously as "supportive" of an hypothesis. It is not until a study has been repeatedly replicated that we begin to view findings as conclusive. Henri correctly points out that there are many properties of computer conferencing that can be studied using content analysis techniques. Rather than restrict the list of possible applications, we are concerned that too many researchers insist upon devising their own techniques for analysis, thereby restricting the replicability of the analysis.

Systematic

The term 'systematic' has two denotations, orderly and systemic. Reber (1996) defines 'systemic' as "... a more or less well structured set of ideas, assumptions, concepts and interpretative tendencies which serves to structure the data of an area..." (p. 780). Kanuka & Anderson's (1998) study provides a good model of a systematic study. At the outset of their investigation, the authors recognized an association between the attributes of computer conferencing and the tenets of constructivism. Therefore, they began by identifying their perspective as constructivist and then selected a transcript analysis instrument that views communicative behavior in terms of active, collaborative, construction of knowledge. Other studies, however, combine irreconcilable paradigms in their analysis of data. Howell-Richardson & Mellor (1997) identify this issue in Henri's (1991) classification schema, noting that "...the level of description at the social, cognitive skills or interactivity levels was dependent on a mixture of theoretical approaches, which were not necessarily mutually consistent" (p. 69). As late as 1998, Bullen was casting a wide net, alternately sampling from Ennis' cognitive perspective of critical thinking, Henri's behaviorist perspective of interactivity,

and Harasim's constructivist perspective of participation. The exploratory nature of these studies reflects the immaturity of the field rather than the deficiencies of the methodology or the naivete of the researchers.

Quantitative

The third attribute that Berelson offers is 'quantitative.' The impetus to perform a content analysis often comes from an instructor's intuitive impression that something valuable has occurred in a conference, however, the scientific investigation of this 'impression' should proceed in an empirically rigorous manner whether the research team chooses quantitative or qualitative methods.

We see considerable value in qualitative approaches, especially in the early phases of any research topic. Qualitative analyses of conferencing transcripts often take the form of frequency counts of the number of messages followed by identification and categorization of major themes that emerge from the transcript data. Often researchers follow "grounded theory" methodology (Corbin and Strauss, 1990) when identifying, describing and verifying these categories.

Transcripts can also be "read" in a qualitative manner to support reflection and induce individual learning. This is especially useful in participative or action research, but such work falls outside of the traditional definitions of content analysis as defined by Berelson. Many studies compliment a quantitative analysis with additional qualitative research such as open ended survey's and interviews (Bullen, 1998; Kanuka & Anderson, 1998; Fieberg & Kramer, 1998; Weiss & Morrison; 1998; Kanuka & Anderson, 1998). The combination of methodologies has been used to "...provide a more comprehensive picture of on-line discussion (Angeli, Bonk, & Hara, 1998, p. 8); triangulates results; and uncover issues that the researcher may overlook from her point of view.

However, we also believe there is a need to establish techniques and that can be used across contexts to produce comparable results in which the inevitable error and variation induced by the researcher and their methodology is minimized. The distinction between quantitative and qualitative research can be obscured as reflected in Angeli, Bonk, & Hara's statement that "...Henri's [1992]

model for content analysis ... was used to qualitatively analyze the electronic discourse" (p. 2). A review of their study shows that, in fact, they used Henri's model to *quantitatively* analyze the *quality* of learning in the conference.

Descriptive

Berelson (1952) characterizes content analysis as primarily a descriptive technique; Riffe, Lacy, & Fico (1999) call descriptive studies "reality checks" (p.#). Like all new educational technologies, computer conferencing was introduced amidst glowing hyperbole. Experienced educators and astute theorists wrote eloquently on the positive attributes of text-based communication, asynchronous dialogue and collaborative learning environments. But, early on researchers such as Beckwith (1989) and Mason (1991) identified the need to empirically test these claims. 10 of the 12 studies we reviewed were descriptive, meaning they described, organized, and summarized what was occurring in a specific computer conference (Bullen; 1998; Kanuka & Anderson; 1998; McDonald; 1998; Weis & Morrison, 1998; Angeli, Bonk, & Hara, 1998; Mower, 1996; Zhu, 1996; Webb, Newman, & Cochrane; 1995; Henri, 1991). Important information has been collected on several of the themes mentioned above which gives subsequent researchers a foundation upon which they can build. For example, Bullin (1998) characterized participation in his group as "low to medium" relative to participation levels in Harasim's (1993) study (p. 7). These studies provide a rich source of anecdotal data and a model for the acquisition of more, fundamental information.

Inferential

Often researchers want to extend the purpose of content analysis from simple description to inferential hypothesis testing. Borg & Gall (1989) discuss this shift in the context of educational research: "Whereas most early studies employing content analysis relied on simple frequency counts of objective variables (e.g., spelling errors), recent studies more often aim at using content analysis to gain insights into complex social and psychological variables" (p. 521). To this, they add the following caveat: "Such studies are much more difficult to

carry out than the simple frequency studies and often depend on a researcher's high level of sophistication ..." (p. 521). Ahern, Peck, & Laycock's study of 1992 was the first to combine random assignment to groups and controlled manipulation of variables with the content analysis technique in the analysis of computer conferencing transcripts. This approach was advanced by Howell-Richardson & Mellar (1996) and Marttunen (1997) who were able to make convincing conclusions concerning different conditions.

We are concerned when researchers use basically qualitative techniques (with no or inadequate discussion of reliability) and then use results of these qualitative measures to generate inferential statistics. For example, Blanchette (1999) scores messages using a self-developed coding format, presents no data related to reliability of the instrument and then uses the information to infer significant differences in gender and age of the participants.

Content Issues

Manifest content

Berelson, Holsti (1969), and Riffe, Fico, & Lacy (1998) concur that "...the requirements of scientific objectivity dictate that coding be restricted to manifest content ..." (Holsti, p. 12). Seven important conferencing issues have been studied in this manner including, participation (Fahy et. al., 1999; Angeli, Bonk, & Hara, 1998; Bullen, 1998; McDonald, 1998; Howell-Richardson & Mellar, 1996; Henri, 1992), interaction (Fahy et. al., 1999; Angeli, Bonk, & Hara, 1998; McDonald, 1998; Mower, 1998; Howell-Richardson & Mellar, 1996; Zhu, 1996; Ahern, Peck, & Laycock, 1992; Henri, 1992); group dynamics (McDonald, 1998; Howell-Richardson & Mellar; 1996), types of teacher discourse (Howell-Richardson & Mellar, 1996; Ahern, Peck, & Laycock, 1992) levels of argumentation (Marttunen, 1997); emotions (Weiss & Morrison, 1998); and linguistic variation (Blanchette, 1999) (see Table 1). Doubless, there are other manifest behaviors of interest to scholars of CMC interactions that will be measured and described in future study. Manifest content lends itself most readily to formal content analysis and early descriptive accounts of computer conferencing have helped us to describe and

quantify typical patterns of interaction, discourse and participation - each useful in understanding the media's use in formal education contexts.

Latent Content

Not all research questions, and especially many of the most interesting ones, can be answered by focusing on the surface content of the transcripts. As Colford (1996) points out "once we have written something down, it can remain there for others to see as a lasting record of our thoughts. But what also remains there for others to see is the person, the self, the interior being, responsible for our thoughts". (p. 40). The challenge for researchers lies in developing methodologies that allow us to decipher and understand these latent processes while keeping our methods transparent and manifest. The overriding concern of many educational researchers is whether or not computer conferencing can facilitate higher-order learning outcomes. Educational theorists are coming to regard 'higher-order learning outcomes' not as overt products, but rather as covert processes (Anderson, ##, Garrison, Archer & Anderson, ### Rasmusen ####. Researchers have proposed some interesting strategies for measuring latent content in conferencing transcripts.

In the studies that we reviewed, cognitive processes were the most commonly investigated latent variable. Henri's (1991) and Zhu's (1996) classification schemata look for "cognitive' dimensions" in the transcripts. Other's, beginning with Mason (1991), look for evidence of 'critical thinking' in its various definitions (Fahy et. al., 1999; Bullen, 1998; Newman, Web, & Cochrane, 1996).

Experienced content analysts argue that measuring latent content is inherently subjective and interpretative. Henri's taxonomy has been criticized on these grounds by Angeli, Bonk, & Hara (1998), Newman, Webb, & Cochrane (1996), and Howell-Richardson & Mellor (1996). Newman, Webb, & Cochrane's coding protocol that accompanies their instrument clearly illustrates the practical problems of identifying latent variables:

Rather than classify every statement in a transcript as, e.g.

critical assessment or uncritical acceptance, we mark and count the obvious examples, and ignore the intermediate shades of grey. This eases the task of the assessors, since there is less need for subtle, subjective, borderline judgements...Of course, one statement might show more than one indicator, ... Or indicators can even overlap (p. 69).

The implications of this protocol on objectivity and reliability are apparent.

Instead of identifying latent variables at the coding level, Riffe, Lacy, & Fico (1998) suggest postponing this type of analysis to the interpretive stage, "...at which time," Holsti (1969) comments, "...the investigator is free to use all of his powers of imagination and intuition to draw meaningful conclusions from the data (pp. 12-13). Two studies have taken this approach. To begin her study, Mason (1991) induced a typology of common communicative behaviors in conferencing transcripts. Her typology includes, use of personal experiences related to course themes; reference to appropriate material outside the course package; comments on others' opinions, both student and tutors; introduction of new issues for discussion; summaries of previous messages; students posing questions for the group; and tutors acting as facilitators (p. 168). Mason and Weiss & Morrison (1998) used these manifest elements to code the transcripts. Then, in the final stages of their study, they proposed an association between the manifest behaviors and latent variables such as critical thinking, judgement, and initiative.

A more popular alternative has been to reverse this process. Bullen (1998), McDonald (1998), Gunawardena, Lowe, & Anderson (1997), Marttunen (1997), Newman, Webb, & Cochrane (1997), Zhu (1996) and Henri (1991) chose to define the latent variables and then deduce communicative behavioral indicators of these variables. For example, Henri's 'surface processing' category was identified in the transcript by behaviors such as "repeating what has been said without adding any new elements" (p. 130). Both of these approaches, inductive and deductive, have been useful for studying latent variables through a survey of manifest content.

Unit of Analysis

Part of conducting a quantitative study involves identifying the units that will be classified and recorded. In content analysis nomenclature, this process is called 'unitizing' (Krippendorff, p.#). Researchers have experimented with different types of recording units with varying degrees of success. Their goal has been to select a unit that multiple coders can objectively identify, and a one that exhaustively and exclusively encompasses the sought-after construct. The research that we reviewed points to a frustrating, negative correlation between these two criteria. Fixed units such as single words or entire messages are objectively recognizable, but they do not always properly encompass the construct under investigation. Dynamic units such as Henri's 'unit of meaning' properly delimit the construct, but invite subjective and unreliable identification of the unit..

Sentence

Units such as the word or the sentence are called 'syntactical units' because they are delimited by syntactical criteria. Fahy et. al. used the sentence as their recording unit to help meet their goal of developing an instrument that is easy to use and reliable. During a preliminary analysis, they reported interjudge agreement figures as high as 94%. Our experience with this unit of analysis was less encouraging. The objectivity of a syntactical unit is confounded by the idiosyncratic nature of conferencing communications. The syntax in the conferences we studied combined the telegraphic style of email with the informality of oral conversation. The following selection from one of our transcripts is typical:

Certain subjects could be called training subjects...i.e. How to apply artificial respiration....as in first aid...and though you may want to be a guide on the side....one must know the correct procedures in order to teach competency...other subjects lead themselves very well to exploration and comment/research.

It is obvious that the strength of the sentence unit--reliable identification--did not materialize in this example. It also introduces an additional subjective step to the research process in which coders must first interpret the utterances

and perform a syntactical transformation. Also, sentence level coding yields an enormous amount of cases. In a 13 week conference we identified over 2,000 sentences.

Paragraph

Angeli, Bonk & Hara (1998) attempted to use a slightly larger syntactical unit, the paragraph. This unit, by definition, could significantly reduce the number of cases. However, as the size of the unit expands, so does the likelihood that the unit will encompass multiple variables. Or, conversely, that one variable will span multiple paragraphs. Our experience did not support the authors' optimism that "...college-level students should be able to break down the messages into paragraphs" (p. 9). Often, a full line of space or a tab was used for purposes other than delimiting a single coherent and unified idea accompanied by a group of supporting sentences. And, once the syntactical criteria are lost, the definition of the unit of analysis as 'paragraph' becomes meaningless, so what the coders are identifying are, in fact, graphical blocks of text. Angeli, Bonk, & Hara's ad hoc coding protocol reveals these problems: "...when two continuous paragraphs dealt with the same ideas, they were each counted as a separate unit. And when one paragraph contained two ideas, it was counted as a two separate units" (p. 9). Using this protocol, Angeli, Bonk & Hara settled for an aggregate interjudge agreement figure of 74.6 % which was "...deemed adequate given the subjectiveness of such scoring criteria" (p. 9).

Message Unit

Marttuunen (1997) looked for levels of argumentation and counterargumentation in transcripts. Defining 'argument' as 'a supported proposition,' Marttuunen, like Ahern, Peck, & Laycock (1992), used the message as the unit of analysis. This unit has important advantages. First, it is objectively identifiable: unlike other units of analysis, multiple raters can agree perfectly on the total number of cases. Second, it produces a manageable set of cases. Marttuunen and Ahern, Peck, & Laycock recorded a total of 545 and 185 messages respectively, a total that would have been considerably larger if the messages had been subdivided. Third, it exhaustively and exclusively contained the object of

Marttunen's study. Fourth, it is a unit whose parameters are determined by the author of the message. In an exemplary discussion of interrater reliability, Marttunen reported a "...reliability (r)..." of 0.71 ($p < 0.01$), while Ahern, Peck, & Laycock reported interjudge agreement at "...over 90%" (p. 298) when using the message unit as the unit of analysis.

Meaning Unit

Authoritatively and apriorily fixing the size of the unit based on criteria that are tangential to the construct under study was rejected by Henri. Instead, she opted for a phenomenological approach. Quoting from Muchielli, Henri argued that "...it is absolutely useless to wonder if it is the word, the proposition, the sentence or the paragraph which is the proper unit of meaning, for the unit of meaning is lodged in meaning (p. 134; for full citation see Muchielli, 1984; p. 32). The task of explaining what this enigmatic statement meant to pragmatic researchers was taken up by Howell-Richardson & Mellar (1996). Drawing on speech-act theory, they explained that transcripts should be viewed with the following question in mind: What is the purpose of this particular utterance? A change in purpose sets the parameters for the unit. The authors also evaded some of the difficulties that Henri's scheme presents by sticking to manifest content such as the linguistic properties of the post and audience to whom it was directed. Coding a complex, latent construct such as 'in-depth processing' with a volatile unit such as Henri's 'meaning unit' creates large opportunity for subjective ratings and low reliability. Not surprisingly, Henri offers no reliability discussion. Howell-Richardson & Mellar's method has advantages, however, rather than reporting interrater reliability figures, the authors submit the following tantalizing discussion:

Our procedure overcomes both the problem of relying on potentially inconsistent judgements in deciding whether or not a set of wordings constitute a single meaning or more than one and the problem of suggesting that graphic boundaries of the message can be equated with a single communicative act (p. 52).

The selection of the unit of analysis is complex and challenging for

the quantitative content analysis researcher. In our own research we often code transcripts using a selection of units and choose the one that seems to offer the meaningful interpretation while being relatively easy to identify and that allows us to maintain adequate levels of reliability. We also tend to recode the same transcripts using different units of analysis depending upon the nature of the investigation.

Software to Aid Content Analysis

The existence of machine-readable data (the conference transcripts) does not guarantee that the transcripts are available in a format that can easily be analyzed. A first problem is gathering the data in a single text file that contains the entire sampling unit. Some conferencing software does not support export of the complete conference or portions, but rather forces researchers to tediously cut and paste each individual message from a separate window into a larger text file.

Once in a text file, there are a number of software packages that can be used to assist in the analysis process. The most useful are qualitative analysis packages such as Atlas/ti, Nudist and HyperQual. These packages allow the researcher to identify the unit of analysis in the transcript and assign the text to a coding category that has been theoretically defined a-priori or to one that emerges from the analysis process. Later analysis can combine or sort codes into families for more meaningful discussion, presentation or analysis. These packages allow multiple coding of individual passages for use when more than one construct is being investigated and allow multiple coders to work on a single coding task while maintaining identification of the coder for calculation of reliability. A wide variety of reports can be generated from these packages including list and frequency counts of codes with or without the illustrative quotations from the text.

In addition to the hand coding by researchers, many of these packages allow coding to be automated based upon multi-string text search and pattern matching. Other quantitative data can be generated including number of sentences, coding results by individual posters and counts of results from

multiple documents.

Although many of these packages perform major tasks in very comparable ways, we have chosen to use the German product ATLAS/ti (<http://www.atlasti.de>) in our research. This product seems to us to be relatively easy to use, but still powerful enough to perform all of the functions we find necessary for coding of transcripts. We also like the concept mapping tool and export to XML features found in later versions of the programs.

Once the content of the transcripts had been coded and categorized using Atlas-ti, SPSS can be employed for more quantitative and reliability analysis. Preliminary interjudge agreement calculations are performed using SPSS's chi square function, and final interrater reliability figures are calculated with SPSS's Cohen's kappa feature. Cohen's kappa does

Ethics

We conclude with a brief discussion of the issues related to ethical issues related to content analysis of computer conferencing transcripts. Questions of ethical approval and informed consent are important to all researchers. We have had personal experience in which a proposed study was funded and then aborted due to the reluctance of a single individual to allow external researchers to review and read the contents of the computer conference transcript. Alternatively, we have been involved in the tedious process of obtaining ethical clearance from a university ethics approval board for transcript analysis and been left wondering if such approval was both useful or necessary.

Our experience as researchers in a Canadian University operating under ethical approval guidelines set by our University and recommended by Canadian federal research granting councils is probably similar to researchers operating under other jurisdictions, though care should be taken to investigate policy and practice in other domains. Ethical guidelines have been established to protect human subjects from harm as result of participation in scientific investigation. The three Canadian federal granting councils released a Code of Ethical Conduct for Research Involving Humans in 1994. The Code cites four principles to guide

researchers in construction and evaluation of research protocols. These principles are "respect for persons", "non-maleficence", "beneficence" and "justice".

The "respect for persons" principle is grounded upon the right of participants to make "informed choice" as to degree (if any) of participation in the study. This is the area of greatest issue to many researchers. This code defines research participants as "living individuals or groups of living individuals about whom a scholar conducting research obtains (1) data through intervention or interaction with the individual or group, or (2) identifiable private information". Distinguishing between active "action research" in which the researcher takes part in the conference under investigation and one on which the researcher merely examines the subsequent transcript, changes the nature of the "intervention or interaction" between researcher and research subject. We argue that a researcher analyzing the transcripts of a conference, without participating has not intervened in the process and thus has not placed the authors in the position of "research participants". However, the second criteria is relevant in that often transcripts contain "private information" that has been posted to the conferencing group.

Two solutions to this problem are possible. The researcher can request that each participant sign a normal "informed permission" release form in which the standard information is provided to participants including: nature of the investigation, potential harm and benefits are outlined, use of the information and contact information to discuss concerns with the researchers etc. This standard process of subject permission is complicated in a formal education context in which protection of privacy may preclude the release of addresses of students to which the researcher can post release forms. In our experience, sending such forms by email or posting within an administration section of the computer conference, results in the majority of students responding positively to the request, none objecting, but a few not replying at all to the request for participation. In worst case a negative response, or lack of any response forces researchers to either abandon this sample group or to have the postings of individuals who have given permission removed from the transcript prior to

analysis. Removal of individual non-participating posting is possible using search and delete techniques of the analysis software, but in practice becomes problematic in that often postings contain excerpts and quotations from previous postings, any of which may have been made with non-participating subjects. In addition use of personal names is common and eradicating all references to non-participants can be very time consuming. Further, one could narrowly define removal of a non-participants posting itself as an analysis process requiring permission of the participants. Finally, the removal of one or more person's postings may make understanding of the conference thread impossible and decontextualize subsequent postings.

A second more encompassing solution is to reduce the requirement for informed consent, by applying the two criteria of the "research participant" above and concluding that transcript analysis participants are not, by definition research participants. To make such a conclusion one must address the second stipulation that the researcher not obtain "identifiable private information". The use of "search and replace" features of analysis software is then used to change all personal or login names from headers of postings and within the postings to "subject1, subject 2" etc.

The study of computer conferencing transcripts seems to hold little danger of maleficence, and we believe high potential for beneficence - especially in potential to increase learning efficacy of subsequent conferences. The issues of justice seems not to be of major concern and is normally an issue only when conducting research with specialized target groups based on gender, race or social economic status. Thus, issues of informed consent seem most problematic for transcript analysis researchers. There seems no easy solution to this problem, other than for researchers to expect to expend some considerable energy obtaining consent or stripping non-participant postings or personal identification from the transcripts.

Conclusion

Quantitative content analysis is a statistical technique that continues to evolve as researchers from many disciplines gain experience through its

application. This evolution is reflected in the modification of Berelson's succinct and limited definition proposed in 1952 to the expanded definition proposed by Riffe, Fico, & Lacy in 1998:

Quantitative content analysis is the systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules, and the analysis of relationships involving those values using statistical methods, in order to describe the communication, draw inferences about its meaning, or infer from the communication to its context, both of production and consumption (p. 20).

Educational researchers have recognized content analysis as a powerful tool for the study of computer conferencing transcripts; and together, we are learning how to apply the technique in a manner that yields valuable and valid insights into the nature of teaching and learning in these environments. The features described in Riffe, Lacy & Fico's definition are not accidental features of some content analyses: They are the criteria for quantitative studies using this technique. If researchers are to extract valuable information and compare their findings with those of other researchers, each of these criteria must be met.

A critical test is the responsible reporting of reliability figures. As Riffe, Lacy, & Fico insist "failure to report reliability virtually invalidates whatever usefulness a content study may have" (p. 134). The terminology and some of the techniques of content analysis are being used by qualitative researchers, notably those doing "grounded theory" research. This use of the same nomenclature to describe two different research techniques is confusing and we argue not productive for progress in the field. Although argueably no research tradition owns the words "content analysis", it would be helpful for researchers to describe their work as following classical quantitative approaches to content analysis or the more qualitative techniques of grounded theory. Mixing techniques and criteria at the whim of the researcher or the nature of the content confounds common understanding of these complex problems.

Quantitative content analysis allows replication. The results of a study gain immense credibility, persuasiveness, and value if they are repeatedly replicated in

subsequent studies. Further, replication in diverse contexts with instruments that are stable allows us to compare and contrast findings across learning environments. There has been little replication in this field and no stable, easy to use instruments have evolved for use by subsequent researchers. Newman, Webb, & Cochrane conclude their study with an invitation to other researchers to apply and improve upon their protocol; likewise, Howell-Richardson & Mellar suggest that the validity of their method "...is an empirical question" (p. 53). With this statement, they are inviting others to test their method in practice.

Further descriptive studies are needed to identify the salient elements of asynchronous, text-based computer conferencing. Not all of the original hyperbolic claims have been empirically tested, yet. Does asynchronous communication foster more reflection and careful response composition? Does text-based communication lead to more articulate presentation of arguments? If these claims are supported, then inferential testing will play an important role in defining exactly how to facilitate this potential.

References

Ahern, T., Peck, K. & Laycock, M. (1992). The effects of teacher discourse in computer-mediated discussion. *Journal of Educational Computing Research*, 8(3) 291-309.

Angeli, C., Bonk, C., & Hara, N. (1998). Content analysis of online discussion in an applied educational psychology course. *CRLT Technical Report No. 2-98*. Available [Online]: <http://crlt.indiana.edu/publications/crlt98-2.pdf>

Beckwith, D. (1987). Group problem solving via computer conferencing: The realizable potential. *Canadian Journal of Educational Communication*, 16(2), 89-106.

Blanchette, J. (1999). Register choice: Linguistic variation in an online classroom. *International Journal of Educational Telecommunications*, 5 (2), pp. 127-142.

Borg, W. & Gall, M. (1989). The methods and tools of observational research. In W. Borg & M. Gall (Eds.) *Educational Research: An Introduction (5th ed.)*. London: Longman, 473-530.

Bolter, J. D. (1991). *Writing space: the computer, hypertext and the history of writing*. Hillsdale, N.J.: Erlbaum.

Bullen, M. (1998). Participation and critical thinking in online university distance education. *Journal of Distance Education*, 13(2), 1-32.

Capozzoli, M., McSweeney, L. & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3-23.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13, 3-21.

Colford, I. (1996). Writing in the electronic environment: Electronic text and the future of creativity and knowledge. Occasional Papers Series 59. Halifax: Dalhousie University, School of Library and Information Science.

Fahy, P., Crawford, G., Ally, M., Cookson, P., Kellar, V., & Prosser, F. (1999). The development and testing of a computer conferencing transcript analysis tool. Unpublished Manuscript.

Gunawardena, C., Lowe, C. & Anderson, T. (1997). Analysis of a global on-line debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, 17(4), 395-429.

Henri, F. (1991). Computer conferencing and content analysis. In *Collaborative Learning Through Computer Conferencing: The Najaden Papers* (A. Kaye Ed.). London: Springer-Verlag, 117-136.

Holsti, O. (1969). *Content Analysis for the Social Sciences and Humanities*. Don Mills: Addison-Wesley Publishing Company.

Howell-Richardson, C. & Mellar, H. (1996). A methodology for the analysis of patterns of participation within computer mediated communication courses. *Instructional Science*, 24, 47-69.

Kanuka, H. & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of Distance Education*, 13(1) 57-75.

Marttunen, M. (1997). Electronic mail as a pedagogical delivery system. *Research in Higher Education*, 38(3), 345-363.

Mason, R. (1991). Analyzing computer conferencing interactions. *Computers in Adult Education and Training*, 2(3), 161-173.

McDonald, J. (1998). Interpersonal group dynamics and development in computer conferencing: The rest of the story. In *Wisconsin Distance Education Proceedings Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Available [Online]:
<http://www.mrc.gc.ca/publications/publications.html>

Mower, D. (1996). A content analysis of student/instructor communication via computer conferencing. *Higher Education*, 32, 217-241.

Newman, G. (1994). Evaluating the quality of learning in CSCL. Presented at UK CSCW SIG workshop on CSCL, Heriott-Watt University, Edinburgh, March.

Newman, G., 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: An Electronic Journal for the 21st Century*, 3(2), 56-77. Available [Online]:
<http://www.helsinki.fi/science/optek/1995/n2/newman.tx>

Reber, A. (1995). *Dictionary of Psychology (2nd ed.)*. Toronto: Penguin Books.

Riffe, Lacy, Fico. (1998). *Analyzing Media Messages*

Weiss, R. & Morrison, G. (1998). *Evaluation of a graduate seminar*

conducted by listserve. In *Proceedings of Selected Research and Development Presentations for Educational Communications and Technology (AECT)*.

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 (18th National Convention, Indianapolis)*.

Table 1.

Summary of Computer Mediated Communication Content Analysis Studies

Study	Unit of Analysis	Categories	Reliability	Descriptive/ Inferential
Ahern, Peck, & Laycock (1992)	Fixed*: Message	Interaction Complexity of response	Interjudge agreement	Descriptive Inferential
Angeli, Bonk, & Hara (1998)	Dynamic*: Paragraph (multiple codings/unit)	Participation Interaction Social Cognitive Metacognitive	Intrajudge agreement Coder stability	Descriptive
Blanchette (1999)	Dynamic: Markers of informality	Linguistic variation Participation Themes	No reported	Descriptive
Bullen (1998)	Dynamic: Meaning unit	Participation Critical thinking	Not reported	Descriptive
Fahy et. al. (1999)	Fixed: Sentence	Interaction Participation Critical thinking	Interjudge agreement	Descriptive
Henri (1991)	Dynamic: Meaning unit	Participation Social Interaction Cognitive Metacognitive	Not reported	Descriptive

Howell- Richarson & Mellar (1996)	Dynamic: Illocutionary unit	Participation Illocutionary properties Focus (group/task)	Not reported	Descriptive
Kanuka & Anderson (1998)	Dynamic: Meaning unit	Collaborative knowledge construction	Not reported	Descriptive
Kanuka & Anderson (1997)	Dynamic: Meaning unit	Collaborative knowledge construction	Not reported	Descriptive
Marttunen (1997)	Fixed: Message	Levels of argumentation / counter argumentation	Reliability coefficient	Descriptive and Inferential
McDonald (1998)	Dynamic: Idea unit	Participation Social Interaction Cognitive Metacognitive Group development	Cohen's kappa	Descriptive
Mower (1996)	Dynamic: Message (multiple codings/mess age)	Interaction Topics	Coerced** interjudge agreement	Descriptive
Newman, Webb, & Cochrane	Dynamic: Statement	Critical thinking	Coerced interjudge agreement	Descriptive

(1995)				
Zhu (1997)	Dynamic: Paragraph/sentence	Interaction Participation Participant roles Knowledge construction	Not reported	Descriptive

Note: units of analysis for studies in which participation was described quantitatively are not documented in the table. Routinely, the unit of analysis for this measure are number of words, messages, or both.

*'fixed' refers to units of analysis whose size is authoritatively set by the experimenter; 'dynamic' refers to units whose size is phenomenologically dictated by the construct under investigation

** 'coerced agreement' refers to reliability figures that were obtained through discussion between coders.