An Overview of Research Methods for Assessing Content of Online Discussion Forums

Rose M. Marra
Joi L. Moore
Assistant Professor / Associate Professor
University of Missouri – Columbia
Columbia, MO, USA
rmarra@missouri.edu

SUMMARY

The discussion forum is a significant component of online courses. Instructors and students rely on these asynchronous forums to engage one another in ways that potentially promote critical thinking, meaningful problem solving, and knowledge construction. In spite of the importance of these forums, predominantly used methods for assessing the content and outcomes of these forums has often been limited to frequency counts and other quantitative measures. Only recently have researchers developed protocols for conducting meaningful qualitative analysis of online discussion forums. This paper provides an overview of current methods of quantitative and qualitative research methods and protocols for analyzing the content of online discussion forums. For each analysis method we describe the research method, provide an example of a research study that used that paradigm, describe the types of questions that this method can address, and compare the methods in terms of validity and reliability.

KEYWORDS: *e-learning, discussion forum, discussion board, online learning, online courses analysis*

INTRODUCTION

The United States National Center for Educational Statistics reported that in 1997 - 98, 58% of all post-secondary institutions offered Internet-based courses (NCES, 2002). Further, 82% reported that they had plans to increase their Internet-based offerings over the next three years (NCES, 1999). Although there are debates about the relative merits of online education in comparison to face-to-face, the statistics just cited together with the international trend for individuals to rely on electronically-based and asynchronous modes of communication and the increasingly frequent need for workers to update or acquire new skills seem to provide evidence that that online education will continue. Arguably then, researchers and practitioners must direct attention towards methods of conducting meaningful analysis of a key element of online education – the online discussion forum.

The discussion forum is a significant component of online courses. Instructors and students rely on these asynchronous forums to engage one another in ways that potentially promote critical thinking, meaningful problem solving, and knowledge construction. In spite of the importance of these forums, predominantly used methods for assessing the content and outcomes of these forums has often been limited to frequency counts and other quantitative measures. Only recently have researchers developed protocols for conducting meaningful qualitative analysis of online discussion forums. This paper, which extends and builds upon work started by Mason (1992), provides an overview of current methods of quantitative and qualitative research paradigms for analyzing the content of online discussion forums. For each analysis method we describe the

research method, provide an example of a research study that used that method, describe the types of questions that this method can address, and compare the methods in terms of validity and reliability.

BACKGROUND

In discussing online learning, Harasim (1989) describes interactivity as the most striking characteristic of computer mediated communication (CMC) and the factor with the greatest potential to impact learning. Similarly, Garrison, Anderson, and Archer (2001) describe the importance in online learning of creating a "virtual community of inquiry" which allows learners to construct experiences and knowledge through analysis of the subject matter, questioning, and challenging assumptions. In a face-to-face environment, this kind of reflection is often accomplished via synchronous, interactive discussions and problem-solving sessions. Web-based learning courses do not have these opportunities and thus tend to rely on online discussion forums to create these interactions.

Traditionally, educational communications have been grounded in an oral, face-to-face (F2F) tradition. Garrison et al. (2001) describe such communications as being fast-paced, spontaneous, and often less structured than written communications. When working in groups, participants perceive F2F communication as more effective and satisfying than computer-mediated group work (Olaniran, Savate, & Sorenson, 1996) mainly because F2F groups can more easily spend time on such tasks as clarifying and defining responsibilities (Warkentin, Sayeed, & Hightower, 1997). Further, when implemented in properly moderated discussions, oral communication has been shown to support the development of critical thinking skills (Garrison et al. 2001).

Overall, F2F and asynchronous communications differ in terms of the strategies used to manage the conversation. For instance, Condon and Cech (1996) found that in computer-mediated communications, (CMC) participants who were trying to increase communication efficiency and decrease typing requirements also decreased the use of unneeded elaborative statements and repetitions. This finding confirms Garrison et al.'s (2001) description of written communications (as used in online discussions) as being "leaner" because many of the non-verbal signals present in face-to-face oral communication are missing (e.g., body language, pauses). However, even without the richness of the queues found in F2F discussions, McCreary posits that the value of written communication such as used in online discussion forums comes from the necessity of exactness, organization of thought, and clear expression (1990).

The online forum, as noted by Asbell-Clarke and Foster (2004), being text-based provides a unique opportunity to conduct learning research as these forums have the potential to make thinking and reasoning visible both to participants and researchers. Further, even though many acknowledge the critical role the online forum can play in web-based courses, the empirical evidence to indicate that text-based communication used in computer conferencing can facilitate higher-order and critical thinking is just emerging. For instance, Garrison, Anderson, and Archer (2000) report there is "limited" evidence that CMC can facilitate higher-order thinking. However an earlier body of research is more positive in its description of the potential impact of online forums on meaningful learning. Specifically, these studies indicated that participation in online forums could lead to broader and deeper participation in group activities (Kiesler, Siefel, & McGuire,1984; Pullinger, 1986; Spitzer, 1989, as cited in Mazur, 2004). More recently, Jonassen and Kwon (2001) reported that during group problem-solving activities, the CMC participants produced fewer, but more task-related messages than a F2F group. In addition, the CMC group's decision-making patterns were more sophisticated than that of the F2F group. One reason for the

relatively small number of studies addressing meaningful learning via online discussion forums may be the lack of proven research paradigms in this domain.

An early exception to CMC evaluation models was Henri's work (1992) that examined the quality of online postings by focusing on four dimensions – social (e.g., "I'm feeling great today"), interactive (statements that refer to other postings), metacognitive (statements about reasoning), and cognitive. The cognitive dimension is broken down into five types of reasoning skills: elementary clarification, in-depth clarification, inference, judgment, and strategies. But this portion of the model goes further. Specifically, Henri emphasizes that it is insufficient to simply examine the *skills* demonstrated in message content, but rather one must look for evidence of the level of information processing. Unlike prior schemes, Henri's model defines not only the types of skills and interactions demonstrated in online postings, but also attempts to qualitatively define the nature and content of online interactions that evidence cognitive development and meaningful learning.

DESCRIPTION OF ONLINE DISCUSSION RESEARCH METHODS

The scope of this paper does not allow us to review *all* methods that can be used to analyze online discussion forums, however, we can and have chosen what we consider to be several representative or key methods[1]. Online forum analysis methods can be described using the categories: quantitative – descriptive statistics, quantitative – mapping or sequential flows, and qualitative content analysis.

For examples of a qualitative content analysis methods, we chose the IAM (Gunawardena, Lowe, & Anderson, 1997) and Newman, Webb and Cochrane (1996; Newman, Webb, Johnson, & Cochrane,1997) critical thinking model for this analysis because both models attempt to ascertain the presence of meaningful thinking in an online discussion, and further, neither is focused on content acquisition. Secondly, both models are based upon some of the most rigorous work in this young field -- the IAM being based upon Henri (1992), and the Newman et al. scheme being based upon earlier work from Garrison (1992). Also, the IAM is also a strong choice as it is one of the few CMC models with an existing, albeit small, research base. Given that the most frequently used evaluation methods for online discussion forums focused on quantitative descriptive statistics (Mason, 1992), we also address this research method. Lastly, we address recently emerging research combined quantitative / qualitative method for evaluating online discussion forums – sequential or transaction analysis.

Quantitative – Descriptive Statistics

In her review of evaluation methodologies of computer conferencing, Mason (1992) reports that the most often used evaluation method for online discussion forums was "computer-generated statistics about logons, messages sent and read, levels of participation and number and length of entries" (p. 112).

In order for instructors and researchers to have any potential benefit from these data, they must be readily available. However, according to the edu-tools website (http://www.edutools.info/course/compare/index.jsp) that provides product comparisons of online course delivery systems, many systems offer student tracking features that allow instructors to track student logons and their usage of course materials, however, providing descriptive statistics

---

[1] The reader is notified that this paper is not intended to be a comprehensive treatment of research methods for online discussions. For more methods, the reader is referred to Mazur (2004).

of discussion forum usage is not even listed as a comparison feature. A further search revealed that some tools, such as FirstClass (www.firstclass.com) do produce log files containing data about the activity of individual participants, however their web site does not provide details on the nature of these files. Further, there are some new tools that support more sophisticated types of discussion forums than the standard "threaded" discussion forum. Such forum tools enforce a structure on the nature on participants' contributions (Jonassen and Remidez, 2002) and are often accompanied with more sophisticated tools for analyzing the quantitative and qualitative nature of those contributions. One such product is Knowledge Forum (www.knowledgeforum.com) which provides a scaffolded discussion forum environment. Although it does not produce participant data, an aftermarket firm, OISE (Ontario Institute for Studies in Education), has produced an online "Analytical Toolkit" that supports such statistical analysis.

   In her review of computer tools that can contribute to analyzing online forums, Mazur (2004) briefly describes two internal tracking tools that are available as inexpensive internet downloads, that may have as yet untapped potential for quantitatively documenting on-line talk-in-interaction. Both the AXS tracking utility (http://www.xav.com/scripts/axs) and Active Server Page's (ASP) databasing function (http://www.2enetworx.com/dev/projects/statcountex.asp) can provide rather sophisticated tracking of hits by users on web pages. Although neither of these specifically provides statistics for a discussion forum, data about per student hits on particular discussion forum pages can provide an indirect indication of forum usage (e.g. their reading patterns) even when they are not necessarily posting messages.

   Although built-in tools from popular course management software packages do not seem to readily exist to provide descriptive data, the following sorts of metrics would be potentially useful. Mason (1992) lists the number of messages sent and read, levels of participation (although this phrase is not defined), and the number and length of entries for each discussion forum. The following list attempts to provide a more expansive and defined set of metrics that builds upon Mason's work. The metrics are directed towards "threaded" discussion forums – that is "an online dialog or conversation that takes the form of a series of linked messages. The series is created over time as users read and reply to existing messages. Typically, messages in a given thread share a common subject line and are linked to each other in the order of their creation." (www.wested.org/tie/dlrn/course/glossary.html).

- number of postings per student for overall forum
- number of times each posting is read
- total depth per thread; in Figure 1, the "Facilitating comment on Lewis & KEY posting" has a depth of 2 (the initial posting and replies only to that initial posting. In contrast the "Facilitating for Team 2" thread has a depth of three – the initial posting plus replies that extend two levels beyond the initial posting.
- postings per student listed by "depth" of postings– that is the number of postings at depth level $1 - n$ where n is the number of the most embedded posting. For instance in Figure 1, student R. Key has one posting at level 1 (line number 2) and one posting at level 2 (line 6b).

Title		Poster	Date
1.	Week 6 Hernandez	R. Marra	02/26/01
2.	Re Hernandez	R. Key	02/26/01
3.	Re: Hernandez	M. Barks	02/27/01
4.	Question #2 Hernandez	C. Lewis	02/26
5.	FACilitating comment on Lewis & KEY posting	RMM	02/27/01
	a. RE: Question #2 Hernandez	M Bark	02/27/01

6. Facilitating for Team 2: Question #2 Hernandez Case	S. Hartman,	02/27/01
a. RE: Facilitating for Team 2: Question #2 Hernandez Case	M. Barks	02/28/01
b. RE: Question #2 Hernandez	R. Key	02/28/01
7. Facilitating for team 2	S. Hartman	02/28/01
a. RE: Facilitating for team 2	RMM	03/02/01
i. RE: Facilitating for team 2	B. Bolton	03/02/01
b. RE: Question #2 Hernandez	D. Gilman	03/01/01

Figure 1: Sample illustration of threaded discussion postings.

Thus Mason's "level of participation" might be some combination of number of postings, depth of postings and reading patterns for a particular forum. The actual way an instructor or researcher might define this metric, I would argue, should depend on the purpose of participation in the forum and what type of learning the instructor hopes to promote via forum participation. For instance, if the forum is designed to be used to have a student team create an instructional design solution for a provided case study, then it would be important for all team members to participate meaningfully in the forum. In that case the instructor would most likely want to include number of postings per student as well as an indicator of the depth of postings per student to define "level of participation". If on the other hand, the forum is designed to simply be a place to post questions about course readings or assignments, then "level of participation" may simply be defined in terms of student's reading of that forum.

Qualitative Content Analysis- Interaction Analysis Model

In her discussion of CMC evaluation methods, Mason (1992) noted that "the most obvious data available to conferencing evaluators – the transcript of the conference interactions – is paradoxically the least used" (p. 113). Content analysis refers any process that is a systematic, "replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding" (Krippendorf, 1980; Stemler, 2001).

As mentioned previously, Henri (1992) proposed one of the earliest "content analysis" schemes for online discussion forums. Henri's work was significant as it focused on the actual meaning of individual and combined posting rather than examining descriptive statistics or patterns of postings. Similarly, Garrison et al. (2000) proposed a model for characterizing high-quality online interactions that contained three elements: social presence, teacher presence, and cognitive presence. Gunawardena, Lowe, and Anderson (1997) continued down this path of seeking evidence of knowledge building or cognition in online forums. Their proposed model acknowledges Henri's framework but identifies the model's basis in a teacher-centered learning paradigm as a weakness. In response, Gunawardena et al. proposed a content analysis model based on a constructivist paradigm designed to detect evidence of knowledge construction. Their "Interaction Analysis Model" (IAM) was developed in an attempt to further understand and describe the processes of negotiating meaning and knowledge co-construction in a collaborative online discussion environment (Gunawardena et al., 1997). The researchers used the transcripts of a multi-week online debate to develop a model that posits five phases learners must move through (not necessarily sequentially) as knowledge is being constructed. Each phase is described in Figure 2[2]; phase names are shown in italics and operations associated with that phase follow.

---

[2] Note that the original work further defines each phase. For instance, Phase II included further definitions for sub-phases A – C, and for Phase III there are sub-phases A-E.

Phase I. Sharing/ comparing of information: statement of observation or opinion; agreement between participants.
Phase II. Discovery / exploration of dissonance / inconsistency amongst participants: identifying areas of disagreement; asking and answering questions to clarify disagreement.
Phase III. Negotiation of meaning / knowledge co-construction: negotiating meaning of terms and negotiation of the relative weight to be used for various arguments.
Phase IV. Testing / modification: testing the proposed new knowledge against existing cognitive schema, personal experience or other sources.
Phase V. Phrasing of agreement and applications of newly constructed meaning: summarizing agreement and metacognitive statements that show new knowledge construction.

Figure 2. IAM Phase definitions.

To apply this protocol, coders read each posting and apply the most applicable phase or phases from the IAM. In many cases, coders may work on multiple sentences, or a paragraph or two with a single phase (Gunawardena et al., 1997; Marrra, Moore & Klimczek, 2004). Depending on the nature of the forum, postings may contained multiple paragraphs and address several topics, and thus two or more phases may be applied to a posting. In the IAM coding example in Figure 3, the first two paragraphs provide further detail on points that have already been made in prior postings; thus they represent elaboration on an agreed upon point which is the essence of Phase I (see Figure 2). The last paragraph is Phase III, negotiation of meaning or knowledge co-construction, because the student builds on prior posting to construct the new idea that "Spaulding is not getting the whole picture" because no one has addressed student background skill, class organization, etc.

There might be a counterpoint here.
1. Hernandez is touted as a 'super teacher', and has awards for her graduate experience, she has been out of the classroom for 10 years**.** Having worked with the best equipment and labs in the Aerospace industry, she may be unprepared to determine how this atmosphere can be translated to the classroom
2. While Spaulding is enthusiastic and possesses ISD skills, he is not an expert on engineering principles. If Hernandez is overwhelmed with time concerns, she might not be able to offer Spaulding the assistance he needs to develop the new lab and other course adjustments effectively. **PHI (first 2 paragraphs)**
Also, the 'final solution' is merely a list of lab materials. No one addresses student background skill, class organization, time management, or resource issues, such as putting the grad students to use. Because of this I feel that Spaulding is not getting the 'whole picture'. **PHIII (final paragraph)**

Figure 3. IAM coding example (adapted from Marra, Moore and Klimczek., in press)

Results of coded transcripts are categorized into the five phases to provide a distribution that indicates the degree of knowledge construction activities. Marra et al. (in press) conducted inter-rater reliability checks in their use of the IAM. To do so, they adopted the convention based upon prior research using this protocol (Beaudrie, 2000) to use the most advanced phase from each posting as the basis for inter-rater checks. For example, the posting in Figure 3 would be considered an overall "Phase III" because this is the most advanced phase applied to this posting. Hence, the unit of meaning became the entire posting, rather than phrases, sentences, or paragraphs within the posting. Given the developmental nature of this protocol (Gunawardena et al., 1997) and that the highest levels of advancement of knowledge construction is directly implied

by the most advanced phases, this convention seems both logical and viable. Marra et al. (in press) then resolved a portion of their inter-rater differences following a process based on Chi's (1997) recommended process for resolving discrepancies between coders.

Qualitative Content Analysis – Newman, Webb and Cochrane

The Newman, Webb, and Cochrane (1996; Newman, Webb, Johnson, & Cochrane, 1997) content analysis model is designed to measure critical thinking. This protocol is based on Garrison's (1992) 5-stage critical thinking model: problem 1) identification, 2) definition, 3) exploration, 4) evaluation, and 5) integration. These researchers instantiate indicators of critical thinking via approximately 40 codes in categories such as relevance, justification, novelty, and ambiguities, each with a "+" or "-" appended to indicate whether the coded statement contributes to ("+") or detracts ("-") from critical thinking development. A complete list of the code categories is provided in Figure 4, followed by a sample of the expanded coding categories (e.g. the N – novelty – code includes five subcodes that can either be used as a positive or negative indicator.

Category Complete List: Relevance; Importance, Novelty, Outside knowledge, Ambiguities, Linking, Justification, Critical assessment, Practical utility, Width of understanding.

Category Rating Examples	Positive Indicator	Negative Indicator
R+- Relevance	R+ relevant statements	R- irrelevant statements, diversions
I+- Importance	I+ Important points/issues	I- unimportant, trivial points/issues
N+- Novelty. New info, ideas, solutions	NP+ New problem-related information	NP- Repeating what has been said
	NI+ New ideas for discussion	NI- False or trivial leads
	NS+ New solutions to problems	NS- Accepting first offered solution
	NQ+ Welcoming new ideas	NQ- Squashing, putting down new ideas
	NL+ Learner brings new things in	NL- dragged in by tutor

Figure 4. Subset of Newman, Webb, and Cochrane model codes as taken from Newman et al. (1996).

In contrast to the IAM, where the unit of analysis varies from phrases, sentences or the entire posting, for the Newman, et al (1996) model, researchers are to apply the codes at the sentence or phrase level. Newman et al. (1996) indicate that coders should insert the applicable code(s) at the beginning (e.g., "<O+", applying outside knowledge to the problem) and the end (e.g., "+O>") of the unit of analysis. Figure 5 shows the Newman coding scheme applied to the same posting coded with the IAM in Figure 3. Marra et al. (in press) report using the OC+ code because the student was referring to course material, the case text, to make his or her point. Other codes were applied in a similar manner.

<NP+R+There might be a counterpoint here. <L-OC+ Hernandez is touted as a 'super teacher', and has awards for her graduate experience, she has been out of the classroom for 10 years.+OC>-L> <L+Having worked with the best equip and labs in the Aerospace industry, she may be unprepared to determine how this atmosphere can be translated to the classroom.+L>> <L-OC+ While Spaulding is enthusiastic and possesses ISD skills, he is not an expert on engineering principles.+OC>-L>. <L+If Hernandez is overwhelmed with time concerns, she might not be able to offer Spaulding the assistance he needs to develop the new lab and other course adjustments effectively. .+L>> <OC+Also, the 'final solution' is merely a list of lab materials. +OC> <NS+No one addresses student background skill, class organization, time management, or resource

issues, such as putting the grad students to use. +NS><C+<NI+It is because of this that I feel that Spaulding is not getting the 'whole picture'. +NI>+C>+NP +R>

Figure 5. Passage coded with Newman, Webb and Cochrane protocol.

The Newman et al. (1996) model includes a formula for calculating a "critical thinking ratio" (CT) based on the frequencies of plus (+) or minus (-) codes for each letter category. $CT = (x+ - x-) / (x+ + x-)$, where x is a letter-labeled category such as J or N, followed by either a plus or a minus sign. For each coding category (e.g. "J", justification), one tallies the number of positive (x+ in the above formula) and negative (x-) contributions, and then calculates the ratio shown above. This produces a measure that is independent of the quantity of participation, reflecting only the quality of the messages. Ratios for an individual category may range from a -1 (all uncritical, all surface) to +1 (all critical, all deep). For example, if in an online discussion there were a total of 29 "J+" codes and 11 "J-" codes, the critical thinking ratio for the novelty category would be 0.16.

Quantitative – Mapping / Sequential Analysis

Sequential analysis methods for examining discussion board content are designed to provide a view of discussion participation not offered by either the statistical metrics or content analysis methods. Sequential analysis methods are used to identify patterns in interactions within a domain of interactions (Bakeman & Gottman, 1997) – in this case within discussions. Proponents of sequential analysis argue that content analysis methods do not provide data on the relationships between messages nor how message sequences affect subsequent postings and ultimately the learning outcomes of the discussion (Jeong, 2003).

Jeong (2001) describes a method and software program – Discussion Analysis Tool (DAT) – designed to help researchers examine and measure student interactions and sequences of interactions that may lead to critical thinking outcomes in threaded discussions. DAT computes "transitional probabilities between critical thinking events providing the basis for measuring and describing the relationship between threaded messages and student interaction" (Jeong, 2001, p. 28). An additional feature is that the probabilities generated by DAT can be converted into graphical depictions that provide what may be a useful visual representation of student interactions.

Jeong (2001) bases DAT on the theory of dialogism (Bakhtin, 1981) that views language as a part of a larger social context where potential word meanings interact and possibly impact future meanings or interpretations. This theory proposes that critical thinking and meaning are produced by the relationships of one series of statements to another, and further that meanings are affected by conflict that may occur in the interactions. Although not described in the source material in these terms, this conflict serves to produce cognitive dissonance that can then drive inquiry, reflection and articulation of understandings, evidence and assumptions. Jeong (2001) emphasizes that a key assumption of this theory as operationalized in analyzing online discourse is that one cannot analyze individual postings but rather must look at the relationships and transitions between postings. This is the basis of DAT.

Jeong(2001) defines an "interaction" as a given discussion posting and the responding message. His methodology for using DAT essentially builds on content analysis by applying a coding scheme consisting of twelve categories (position statements, simple agreements or disagreements, arguments, personal experiences, literature, formal data, personal or hypothetical actions and choices, evaluation or critiquing of arguments, summary, negotiation or conclusions, and process

comments) to one of the course debates[3]. The coding results were used as input to DAT where it "identified and followed the links between messages" (Jeong, 2001, p. 30). DAT is capable of producing the following types of results.

- Tables of relative frequencies of one event being following by another. Table 1 is adapted from Jeong (2001) and shows the relative frequencies of one type of message being followed by another type. For instance, a "disagreement" message was followed by an Agreement 38% of the time and by another disagreement 13% of the time. These probabilities are based the total number of disagreements observed and the total number of responses to those initial disagreements. Jeong clearly acknowledges that one must balance one's interpretations of these results given the small cell sizes (e.g. n = 5).
- A transitional state diagram as pictured in Figure 6 is a visual illustration of the flow of events from a discussion where the circles show different "events" or codes from the discussion and the arrows depict the transitional probabilities between codes.

Response: Initial Message Type	Agree	Disagree	# Responses	# Initial message
Disagreement	.38	.13	8	5

Table 1. Sample data produced by DAT – event frequencies.

Using these types of data, researchers are able to combine content analysis results with quantitative methods and perform such analyses as tests of significance (using z-scores) for interaction sequences that occur at higher probabilities than other sequences of interactions.
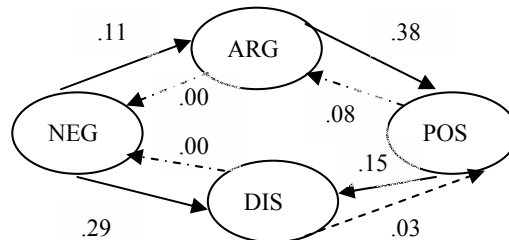

Figure 6. Transitional state diagram example, adapted from Jeong (2001)

METHOD COMPARISONS

Having described a variety of research methods used for analyzing online discussion forum contents, we provide a comparison of the methods in terms of the research questions they are suited for addressing, and their validity and reliability.

Research Questions

The research questions that each method is most suited to address are depicted in Table 2.

Method	Research Questions
Quantitative – Descriptive Statistics	Simple questions on frequencies of postings, readings and thread depth
CA – IAM	What type of knowledge construction is occurring? How much of each type? "Types" are defined by the five phases of the IAM (see

---

[3] The reader is referred to Jeong (2001) for a more complete description of each coding category.

	Figure 2).
CA – Newman protocol	"How much" critical thinking is occurring for each of the codes in the scheme?
Sequential Analysis (DAT)	What sequences of interactions /message flows occur? What sequences occur most frequently?

Table 2. Research questions for each analysis method

The results for the quantitative – descriptive statistics methods are obvious and will not be discussed further. Typical results when using the IAM (Gunawardena, et al, 1997) include a report of the percentages of codings or postings that fell into each of the five phases (e.g. 21% in Phase I, … 5% in Phase V). Researchers are then likely to discuss what overall portion of the postings fell into phases III and above since these are the phases that constitute knowledge construction in some manner. When applied, the Newman, et al (1996) protocol produces a series of CT (critical thinking) ratios for the forty-some codes that are provided in the scheme. According to the scheme, ratios that approach 1 are the highest indicators of critical thinking. Sequential analysis as exemplified by DAT for online discussion boards produces percentage frequencies of event sequences (see Table 1), Z-scores for event sequences which can be used to test significant differences between frequencies of sequences and a visual representation of sequence transitions (see Figure 6).

As noted previously, and as Table 2 shows, the two content analysis methods provide the opportunity to begin to answer more substantive questions regarding the actual "content" of the discussion postings in comparison to the descriptive statistics /metrics method. The relative frequency of their usage in the past few years (Mazur, 2004) is a testament to the popularity of these methods with researchers as they become more interested in questions pertaining to the discussion content evidenced in relationship to, for instance, the discussion task that students are presented with (Marra, et al, 2004).

Further, the method used by DAT, can clearly be combined with most content analysis coding schemes. One could clearly code a discussion using the IAM and then apply DAT's sequential analysis methodology – thus allowing one to answer questions not only about the types of knowledge construction evidenced in the data but also regarding sequences of interactions. This combination of methods would seem to produce the most powerful research paradigm for analyzing both message content as well as the "impact" of message flows on future messages.

Content analysis cannot, as argued by Jeong (2001) to "examine the relationships between threaded messages and how message sequence and group processes affect subsequent discussion and cognitive outcomes" (p. 397). If researchers are interested in understanding how the relationships between messages impact group processes and interactions, then content analysis alone may not be enough .. but rather the type of interaction analysis proposed by Jeong (2001) may be necessary. The authors would agree with this position when comparing Jeong's research method with a content analysis scheme such as the one proposed by Newman, et al (1996). The Newman protocol requires that one code statements in isolation of one another and results in a set of individual frequencies – individual by the very nature of the code definitions. The IAM however, does to some degree take into account the overall flow of postings and their impact on one another as each code is based upon the degree to which knowledge is constructed --- and to assess knowledge construction one must consider each posting in the context of the overall flow of postings. Also, if during coding, researchers applied and used all applicable IAM phases for every single posting, researchers would then be able to address questions concerning the progression of knowledge construction development per posting or per individual student.

Reliability and Validity

   The reliability and validity of these methods should be of great interest to researchers. Given the diverse nature the methods we are addressing, we can expect to see diverse issues concerning reliability and validity for these methods. Not surprisingly, the issues are the most clear-cut for the quantitative descriptive statistics / metrics method. To review, reliability refers to the extent to which a measure produces the same results when applied under the same circumstances. Reliability is an indication of consistency of measures across time. They types of measures we have categorized under descriptive statistics / metrics are by their very nature reliable when counting, and calculations are performed with accuracy and when definitions for concepts such as "depth" are well-understood and applied consistently. To ensure reliability, multiple researchers should simply check calculations and frequencies twice and if possible. In addition to simple counting and arithmetic errors, the greatest possibilities for inconsistencies may occur from individuals' differing decisions about which postings to count (e.g. "do we count the posting that simply said – I didn't mean to submit my prior posting'"). Once such decisions are made, this methodology should be highly reliable. The authors note that the addition of these metrics to commonly used discussion software would not only ensure reliability but also make these metrics more readily available.

   The validity of these metrics is less well defined. Validity refers to whether a measure, in fact, measures what it purports to. For instance, for the Newman, et al (1996) protocol, we discuss whether in fact, this protocol provides measures of critical thinking. For the quantitative descriptive statistics / metrics method, the question of validity comes down to what researchers say that these metrics actually measure. When taken at face value from their statistical definitions, these measures are highly valid. For instance, the tally of the total number postings in a forum is a valid indicator of just that – the total number of forum postings. Similarly unremarkable statements can be made of such metrics as number of postings per participant per forum, average postings per participant, etc. These metrics are not open to much interpretation and as long as researchers do not attempt to attribute more meaning to them than they have, then the metrics are valid. Validity issues may emerge when one attributes more meaning to a metric, beyond the simple definition of the metric. For instance, if one were to tally the per participant posting frequencies for a given forum and then conclude that the top three participants engaged in the most "meaningful" forum participation, then critics would need to call into question the validity of posting frequencies as an indicator of "meaningful" participation. Similarly, it is most likely not valid to say that a forum A with more postings than a forum B is a more "successful" forum – unless of course one is willing to define "successful" based only upon total number of postings. Certainly the metrics that fall under this method category can provide a useful and valid set of parameters to describe certain aspects of the overall forum – however, their validity is questionable in any application that exceeds the simple definition of the metric being considered.

   As with almost any qualitative methodology, the reliability issues associated with both content analysis methods are centered on the ability of more than one rater to consistently apply the coding scheme. As described earlier, for the IAM researchers often apply a single phase for a given posting (Beudrie, 2000; Marra et al, in press). Applying the protocol at the per posting level makes it easier to use an inter-rater reliability procedure. Using a process recommended by Chi (1997), the coders can compare their ratings for each posting, discuss and resolve differences and achieve an inter-rater reliability of 94%. In their use of the IAM, Marra, et al. (in press) reported that one its disadvantages relevant to reliability issues was that the high level definitions of the phases provided by Gunawardena, et al. (1997) and the lack of coding examples made it difficult for coders to decide which codes to apply. The researchers reported that they purposefully contextualized the IAM to their own forum in order to reduce the ambiguity of the scheme. Thus,

the forum in question required students to offer solutions to an instructional design case study, so the researchers jointly agreed upon definitions of each IAM phase that applied to this task.

The reliability of the Newman et al (1996) scheme presents different issues. In contrast to the IAM, the Newman, Webb and Cochrane scheme is straightforward to apply. The Newman et al. codes are very focused in definition. For instance, the JP + and P+ codes are defined as providing proof or examples, and relate possible solutions to familiar situations respectively. Although one can find *some* ambiguity in almost any statement, relative to the IAM broad phase definitions, these codes' definitions are specific and easy to understand. However, Marra et al. (in press) point to the necessity of choosing amongst the *large* number of potentially relevant codes to apply to a text passage as being a significant reliability issue. As shown in Figure 3, there are approximately 40 codes available for the scheme and given that the codes are intended to be used at the phrase, sentence or paragraph level, it is possible to use many codes for a single posting. Lastly, Marra et al. (in press) reported that the assumptions regarding a standardized unit of analysis and mutually exclusive categories for inter-rater reliability algorithms (neither of which apply to the Newman et al. model) (Stemler, 2001), it was not possible to compute an agreement coefficient for the Newman model results.

The question of IAM's validity is simply a question of whether the results produced when the protocol is applied reliably are in fact an indication of the degree to which knowledge construction occurred in the forum. The IAM was developed through a grounded theory process by which the researchers, with a thoroughly explained theoretical basis from social constructivism (Jonassen, Mayes and McAleese, 1993; Pea, 1993; Vygotsky, 1978), applied their knowledge of these theories to an extensive discussion forum with the purpose of attempting to develop a framework for describing knowledge construction via a coding system. Through their preparations they had made themselves aware of the processes that theoreticians had hypothesized occurred during the knowledge construction process – and they then applied this knowledge to the forum looking in particular for ways of describing how the process of "negotiation of meaning" occurred. Although we do not have a definitive "answer" to the question of the IAM's validity, at face value the definitions of the phases taken from Gunawardena et al's (1997) original work, were indeed derived from commonly used descriptions of the types of activities that take place during knowledge construction activities. Additionally, the model was derived via a rigorous qualitative process for developing grounded theory (Strauss, 1990). Lastly, numerous researchers (Beaudrie, 2000; Marra et al., 2004; Jeong, 2001) have successfully applied the model and provided support for the validity of this grounded theory protocol.

The validity of the Newman et al. (1996) protocol is immediately called into question given the concerns previously raised about applying the protocol reliably. Although the CT ratios are easy to calculate, one must question how to meaningfully interpret them. Although a researcher can purportedly tell if critical thinking activities were evidenced by CT ratio sign and magnitude (-1.0 to +1.0), one is still left with a disparate set of ratios that is difficult to interpret for several reasons. First, the variability in the unit of analysis calls into question the meaning of a set of CT ratios. For instance, what does it mean if 70% of the ratios are positive for a discussion forum? Referring to the methods Newman et al. (1996) recommend for applying the protocol, one cannot conclude that 70% of the analyzed text contributed to critical thinking as each ratio is based on coding frequencies that may refer to a phrase, sentence, paragraph or even an entire posting. Additionally, CT ratios are difficult to interpret because of the challenge to consistently use all the numerous codes and also the problem of variability in coding density (e.g., should one use all 8 codes that apply, or only the most obvious?). Both of these procedural difficulties bring the validity of the coding frequencies and the resulting CT ratios into some question.

For DAT, the reliability can be assumed to be nearly perfect – assuming that the underlying programming is sound and that the same data is provided as input for the tool. The validity of DAT is directly tied to the validity of the coding scheme that provides the data on which DAT operates. As the saying goes (at least in the United States), "garbage in, garbage out". If the coding of the discussion forum is not conducted using a reliable and valid coding scheme, then the interaction sequences and other data produced by DAT will of course not be valid.

CONCLUSIONS AND RECOMMENDATIONS

Discussion board forums are a critical component of online courses and to date researchers have generally limited their analyses of these forums to quantitative reports of frequencies. As other methods of coding forum content emerge it is important for researchers to understand when and how these methods can be used, and the strengths and weaknesses of these methods. This paper has provided an overview of current methods of quantitative and qualitative research paradigms for analyzing the content of online discussion forums.

As summarized in Table 2, each method is appropriate for addressing different types of research questions. Further, each method offers its own unique strengths and weaknesses. On one end of the spectrum, one has the option of using quantitative descriptive statistics and metrics to describe discussion board forum participation. This method has the advantage of being easy to implement and highly reliable – however, it is somewhat limited in terms of the types of research questions it can address. On the other end, one has sequential analysis methods such as DAT (Jeong, 2001) – which builds on content analysis methods and provides quantitative data on discussion interaction sequences. Although the results from the software program DAT are completely reliable and the method has the potential to answer more complex questions regarding discussion boards, the method implicitly inherits all the reliability and validity issues of the content analysis protocol that it uses to do the initial coding.

Given these differing advantages and disadvantages, researchers need to choose an analysis methods based upon what questions they wish to answer, and their ability to carry out the chosen methodology in the fashion intended (e.g. does my research team have the qualitative methods expertise to conduct a content analysis?). Lastly research questions should be in alignment with the learning outcomes or objectives of the discussion forum itself. Although the relative power of the IAM may be appealing, does it make sense to look for evidence of knowledge construction in a discussion forum whose purpose is to post examples of student work? In this case, simple quantitative metrics might be best.

The methods described in this paper provide researchers with an effective array of choices for answering questions about a critical aspect of online courses – the online discussion forum. Although none of the methods are "perfect" in terms of their applicability to all situations (nor should they be, we would argue) or in terms of their validity and reliability, all the methods reviewed – when applied appropriately -- can provide researchers with effective tools for describing online discussions. If such research is indeed carried out, the community of scholars examining online learning can look forward to an effective research agenda that will ultimately positively impact our abilities to design and implement effective online learning experiences.

BIBLIOGRAPHY

Asbell-Clarke, J. and Foster, J. (in press). *Theoretical frameworks for online learning environments*, TERC publications. www.terc.edu.

Bakeman, R. & Gottman, J. (1997). Observing Interaction: An Introduction to Sequential Analysis. Cambridge, MA: Cambridge University Press.

Beaudrie, B. P. (2000). Analysis of group problem-solving tasks in a geometry course for teachers using computer-mediated conferencing. Unpublished Ed.D. thesis, Montana State University, Bozeman.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. The Journal of the Learning Sciences, 6(3), 271 - 315.

Condon, S.L, & Cech, C.G. (1996). Discourse management strategies in face-to-face and computer mediated decision making interaction. Electronic Journal of communication/LA Revue Electronique de Communication, 6(3).

Cumming, S. and Ono, T. (1997). Discourse and grammar. In T. Van Dijk, (Ed). Discourse as structure and process, pp. 257-291. Thousand Oaks, CA: Sage Publications, Ind.

de Laat, M. (2001). Network and content analysis in an online community discourse. Paper presented at the Computer-Supported Collaborative Learning Conference. Boulder CO: 2001. Retrieved May 10, 2003, from http://newmedia.colorado.edu/cscl/62.pdf.

Duffy, T.M., & Jonassen, D.H. (1992). Constructivism: New Implications for Instructional Technology. In T.M. Duffy & D.H. Jonassen (Eds.), Constructivism and the Technology of Instruction: A Conversation (pp.1 – 16). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Garrison, D. R. (1992). Critical thinking and self-directed learning in adult education: an analysis of responsibility and control issues. Adult Education Quarterly, 42(3), 136 - 148.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: computer conferencing in higher education. The Internet and Higher Education, 2(2-3), 87-105.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. American Journal of Distance Education, 15(1), 7-23.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of global online debate andthe development of an interaction analysis model for examining social construction of knowledge in computer conferencing. Journal of Educational Computing Research, 17(4), 397 –431.

Harasim, L. M. (1989). Online education as a new domain. In R. Mason & A. R. Kaye (Eds.), Mindweave: communication, computers and distance education (pp. 50 - 62). Oxford: Pergamon Press.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), Collaborative Learning Through Computer Conferencing (pp. 117-136). Berlin: Springer-Verlag.

Jeong, A. (2001). Sequential Analysis of Group Interaction and Critical Thinking in Online Threaded Discussions. The American Journal of Distance Education, 17(1), 25-43. Retrieved June 21, 2004 from http://bbproject.tripod.com/SequentialAnalysis_Jeong2003.pdf.

Jonassen, D.H., Mayes, T. & McAleese, R. (1993). A Manifesto for a constructivist approach to technology in higher education. In T. Duffy, J. Lowycky, & D. Jonassen (Eds.) Designing Environments for Constructivist Learning. Berlin: Springer-Verlag.

Jonassen, D.H. & Kwon, H. (2001). Communication patterns in computer mediated versus face-to-face group problem solving. Educational Technology Research and Development, 49(1), 1042 – 1629.

Jonassen, D.H., & Remidez, H. (2002, January). *Mapping alternative discourse structures onto computer conferences.* Paper presented at the annual meeting of Computer Support for Collaborative Learning, Boulder, CO.

Kiesler, S., Siefel, J. and McGuire, T. (1984). Social psychological aspects of computer-mediated communication. American Psychologist, 39, 1123-1134.

Krippendorf, K. (1980). Content Analysis: An Introduction to Its Methodology. Vol. 5. The Sage COMMTEXT Series, Beverly Hills: Sage Publications

Marra, R.M., Moore, J. & Klimczek, A. (in press) "A Comparative Analysis of Content Analysis Protocols for Online Discussion Forums". <u>Educational Technology Research and Development.</u>

Mason, R. (1992). Evaluation methodologies for computer conferencing applications. In A. R. Kaye (Ed.), <u>Collaborative Learning Through Computer Conferencing</u> (pp. 105 - 116). Berlin: Springer-Verlag.

Mazur, J. (2004). Conversation analysis for educational technologists: theoretical and methodological issues for researching the structures, processes and meaning of on-line talk. In D.H. Jonassen (Ed.), <u>Handbook for Research in Educational Communications and Technology, 2nd Edition</u>. Mahwah, NJ: Lawrence Erlbaum Associates.

McCreary, E. K. (1990). Three behavioral models for computer-mediated communication. In L. M. Harasim (Ed.), <u>Online education: perspectives on a new environment</u> (pp. 117 - 130). New York: Praeger.

National Center for Educational Statistics (1999). <u>Distance Education at Postsecondary Education Institutions</u> (Statistical Analysis Report NCES 2000-013): Office of Educational Research and Improvement.

National Center for Educational Statistics (NCES), (2002). Distance Education at Degree-Granting Postsecondary Institutions: 2000–2001, Executive Summary. http://nces.ed.gov/surveys/peqis/publications/2003017/

Newman, D. R., Webb, B., & Cochrane, C. (1996). <u>A content analysis method to measure critical thinking in face-to-face and computer supported group learning</u>. Retrieved July 15, 2003, from http://www.qub.ac.uk/mgt/papers/methods/contpap.html.

Newman, D. R., Johnson, C., Webb, B., & Cochrane, C. (1997). Evaluating the quality of learning in computer supported co-operative learning. <u>Journal of the American Society for Information Science, 48</u>(6), 484-495.

Olaniran, B.A., Savage, G.T., & Sorenson, R.L. (1996). Experimental and experiential approaches to teaching face-to-face and computer-mediated group discussion. <u>Communication Education, 45</u>, 244-259.

Pea, R. (1993). Practices of Distributed Intelligence and Designs for Education. In G. Salomon (Ed.), <u>Distributed Cognitions.</u> Cambridge, MA: Cambridge University Press.

Pullinger, D.J. (1986). Chit-chat to electronic journals. Computer conferencing supports scientific communication. In V. Arms (Ed.) IEEE Transactions of Professional Communications, PC29, 30-33.

Spitzer, M. (1989). Computer conferencing: An emerging technology. In G. Hawisher & S. Selfe (Eds.), <u>Critical perspectives on computers and composition instruction</u>. (pp. 187-199). New York: Teachers College Press.

Stemler, S. (2001). An overview of content analysis. <u>Practical Assessment, Research & Evaluation, 7</u>(17)

Strauss, A. (1990). <u>Basics of Qualitative Research: Grounded Theory Procedures and Techniques.</u> Thousand Oaks, CA: Sage Publications.

Vygotsky, L. (1978). <u>Mind in Society: The Development of Higher Psychological Processes.</u> Cambridge, MA: Harvard University Press.

Warkentin, M.E., Sayeed, L., & Hightower, R. (1997). Virtual teams versus face-to-face teams: An exploratory study of a Web-based conference system. <u>Decision Sciences, 28</u>, 975 – 996.