Volume 8, Number 2, Summer 2009
ISSN: 1541-4914

Replicating the Use of a Cognitive Presence Measurement Tool

Caroline L. Park, RN, PhD
Athabasca University, Canada

Abstract

This paper is a report of the replication of a seminal study on cognitive presence in computer mediated conferencing (CMC) by Garrison, Anderson, and Archer (2001). A comparison of cognitive presence coding by three different researchers is also demonstrated. The study re-ignites debates about what constitutes the segment of CMC data to be coded and the objectivity of this type of data.

Introduction

As research in both distance education and online education advances, more studies attempting to measure educational outcomes are becoming available in the literature. Garrison, Anderson, and Archer (2000b) explain that "the adoption of Computer Mediated Communication (CMC) has far outpaced our understanding of how this medium should be used so as to best promote higher order learning" (p. 1). Jeong (2003) believes that this shortcoming (of few theories and little empirical research relating to CMC) and subsequently "few instructional strategies on student interaction and learning processes in online discussions" occur because there is a "lack of methods and tools capable of measuring group interactions and processes" (pp. 25-26).

In the 1990's Garrison, Anderson, and Archer (2000a, 2000b; 2001) worked collaboratively to develop a Venn model which could explain the learning experience of CMC students, calling it "The Community of Inquiry Model." They then designed a series of studies to test the model. This paper is a description of the replication of one of their studies for which they used a data collection tool to categorize the cognitive level of conference postings in an online university course. Their research question was basically can a coding tool be used "to judge the nature and quality of discourse in a computer conference?"

Replication

This paper stems from a pilot study which had three objectives relevant to the goals of a health studies department in an online university. As a new department with a large database of computer mediated on-line class content, the overall goals of this study related to testing the feasibility of using an archived database for research purposes, using qualitative analysis software, and using the employment of graduate students, operating at a distance, as research assistants. The following is a report of the replication of a seminal report on cognitive presence

in computer mediated conferencing by Garrison, Anderson, and Archer (2001) which was initiated to reach these goals.

Replication research is infrequent, and although it is seen to be important to theory development, funding bodies and publishers prefer to fund and disseminate original work (Park, 2004). This is unfortunate, particularly because qualitative types of research findings (and cognitive thinking can be viewed as qualitative) from one study are difficult to use in a generalized way. When research time and money have been invested in the development of a research tool such as that utilized by Garrison et al (2000a, 2000b; 2001) to measure levels of critical thinking in CMC, it behooves on-line educators to test, use, and further develop the tool so that the findings can be used to improve educational outcomes by improving instructional and technological inputs. In this replication study, the categories of the tool will be substantiated or repudiated, and their definitions and examples embellished. Hendrick (1990) outlined eight criteria to determine whether a replication was exact, partial, conceptual, or systematic. The original study and this replication are very similar in subject characteristics, research histories of subjects, historical context, and physical setting of research. Neither study had a control agent nor were the specific task variables and primary information focus of each similar. A strict replication is attempted.

Table 1

Comparison to Hendrick's Criteria

Hendrick	*Garrison et al*	*This replication*
1. subject characteristics	Graduate students in two courses	Graduate students in one course
2. specific research histories of subjects - prior experiences and or methods used to get them into the study	-not known -volunteers solicited via e-mail	-senior course, students had experience with technology and on-line conferencing - research consent question built into course design
3. historical context – social cultural systemic factors relating to time and place	- web-based course within larger urban bricks and mortar University	-totally online course at distance university
4. general physical setting of the research	-asynchronous online conferencing	-asynchronous online conferencing
5. control agent	-none	-none
6. specific task variables	-web CT management system -two faculty	-web CT management system -one faculty
7. primary information focus	-same presentation of discussion data -same cognitive categories	-same presentation of discussion data - same cognitive categories
8. modes of data reduction in presentation	-Atlas.ti for content analysis - graduate student coders	-Atlas.ti for content analysis - graduate student coders

Review of Original Study

Underlying the Community of Inquiry Model is the assumption that learning occurs through the interaction of cognitive presence, social presence, and teaching presence (Garrison et al., 2001). The construct of interest in this paper is "cognitive presence". Garrison et al. (2001) defined cognitive presence as the construction of meaning through sustained communication. "Cognitive presence is a vital element in critical thinking, a process and outcome that is frequently presented as the ostensible goal of all higher education." (2000a, p. 4). As well, "cognitive presence reflects higher order knowledge acquisition in application and is most associated with literature and research related to critical thinking" (2001, p. 7). Garrison et al. (2000a) continued by stating that, "one of the goals of the broader study, therefore, is to investigate the features of the written language used in computer conferencing that seems to promote the achievement of critical thinking" (p. 7). The Community of Inquiry Model also describes teaching presence and social presence, aspects of CMC which will affect the nature of learning activities and outcomes. However the purpose of this study was to determine if they could judge the nature and quality of the CMC discourse through the use of a data collection tool, regardless of the influence of the other two spheres.

This is a crucial area for academic research because the promotion of critical thinking is an educational goal, regardless of the mode of interaction. The assumption is that critical thinking, as an outcome, is best measured from an individual perspective; that is, as acquired thinking skills and worthwhile knowledge. Judging the quality of critical thinking as an outcome within a specific educational context is the responsibility of the teacher, as a pedagogical and content expert. One advantage that text based communication provides is time for reflection. For this reason, written communication may actually be preferable to oral communication when the objective is higher-order cognitive learning. (2000a, p. 6).

Garrison et al. (2000b) also profess that "from a process perspective, it is assumed that this [the teacher's responsibility to judge the quality of critical thinking] could be greatly assisted by a tool to assess critical discourse and reflection, for the purpose of acquiring intended and worthwhile learning outcome" (p. 3). Their team developed such a tool. This tool is discussed in some detail in the following paragraphs.

In replicating the use of the Garrison et al. (2001) tool, one must assume that "critical thinking or inquiry is seen here as a holistic multifaceted process associated with a triggering event. This triggering event is followed by perception, deliberation, conception, and warranted action" (2000a, p. 19). The sub-categories in the tool are called: triggering event, exploration, integration, and resolution. The assumption is made that they are a hierarchy, with resolution being the highest level of critical thinking.

The tool under discussion is a "taxonomy of expression" or interaction, as identified from CMC. The expression exists as words in postings to conference forums and is segmented into the three major presences: cognitive, social and teaching. Each of the major presences is then further subdivided. The sub-categories of the social and teaching presence areas appear to be non-hierarchical.

Table 2
Community of Inquiry Categories

Presence	Sub-Category	Example
Cognitive presence	Triggering event	Sense of puzzlement
	Exploration	Information exchange
	Integration	Connecting ideas
	Resolution	Applying new ideas
Social presence	Emotional expression	Emoticons
	Open communication	Risk-free expression
	Group cohesion	Encouraging collaboration
Teaching presence	Instructional management	Defining and initiating discussion topics
	Building understanding	Sharing personal meaning
	Direct instruction	Focusing discussion

The description of the sub-categories of cognitive presence, are summarized below from Garrison et al. (2000a).

Triggering event
This is an issue, dilemma or problem is identified or recognized that emerges from experience. Expectations or tasks provided by the teacher often become triggering events. A critical role of the teacher is to shape and, in some cases, discard triggering events. Examples of conference content demonstrating this category are: presenting background information that culminates in a question, and asking a question and posting messages to take the discussion in a new direction.

Exploration
Exploration can be defined as shifting between the private, reflective world of the individual and the social expression of ideas. Early on in this sub-category, students are required to perceive or grasp the nature of the problem and then move to a fuller explanation of relevant information. This is aided by moving between the private and shared worlds – that is, between critical reflection and discourse. Brainstorming, questioning and the exchange of information occur. Examples of conference content demonstrating this sub-category are: unsubstantiated contradiction of previous ideas; many different ideas or themes presented in one message; personal narratives, descriptions or facts [not used as evidence to support a conclusion]; author explicitly characterizes message as exploration, e.g. “does that seem about right? Am I way off the mark?” Participant adds to establish points but does not systematically defend, justify or develop an addition; and offers unsupported opinions.

Integration
Integration is the construction of meaning from the ideas generated in the exploratory stage. This must be inferred from communication. It is the teacher who must diagnose misconceptions; provide probing questions, comments, add additional information and effort to insure continuing cognitive development and to model the critical thinking process. Examples of conference content demonstrating this sub-category are: reference to previous message followed by the statement, e.g. “I agree because..”; building on, adding to others ideas; justified, developed,

defensible, yet tentative hypotheses; integrating information from various sources; textbook, articles, personal experience; explicit characterization of message as a solution by participant.

Resolution

Resolution involves testing ideas/hypotheses, and treatment of content from a critical perspective. Progression to this sub-category/stage requires clear expectations. This may mean moving onto a new problem with the assumption that students have acquired a useful piece of knowledge. It usually entails a vicarious test through thought experiments and consensus building. The process of apparent skipping of phases or making conceptual leaps, are examples of intuition and insight. Examples of conference content demonstrating this sub-category are: vicarious application to real world; testing solutions and defending solutions.

Research Questions

a) Can the Cognitive Presence Tool, developed by Garrison, Anderson and Archer, be successfully applied to different data by a different research team?
b) Can the process of data collection and analysis implemented by Garrison, Anderson and Archer be replicated by a different research team?
c) What degree of replication can be achieved?

Methods

Data Collection

The data which were subjected to interpretation using the Cognitive Presence Tool were compiled from written conference postings in a graduate nursing course. The course is structured as a debate. Twelve students were enrolled. They debated one issue per week, with pairs of students presenting the pro and con arguments and all others adding to the discussion as the week progressed. Three of the debates, those with the largest numbers of postings, were selected for this study. One debate, actually the shortest of the chosen three, was used as practice data.

Two graduate students were hired as Research Assistants (RAs) to do the actual coding. A teleconference with the RAs, who had been provided with the codebook and test debate data via e-mail, was used for training and orientation. Examples were taken from the test data to use in the third column of the code book. The students and the PI spoke openly about the data and the codes. Samples were coded and compared.

Ethics Review

All twelve graduate students agreed that their participation in the course could be used as research data. At the end of each graduate course in this program we asked each student if they would agree to have their input into the course used for research purposes. They were assured that it would be stripped of names. In this course section, all students completed the consent question in the affirmative. As well, the overall proposal received ethical approval from the University Research Ethics Board.

Tool

For this study, the four sub-categories outlined and described in the original study: trigger event, exploration, integration and resolution, were placed in the left hand column of a three column codebook, or tool. In the middle column all potential examples of each of those sections which could be extracted form the description of the original study were listed, with a dictionary definition of those that might have multiple meanings to the coder. In the third column is an actual example of each, which had been selected from the test debate data by the PI of this study. Garrison et al. (2001) do not provide much detail on the "nitty gritty" of the coding interpretations. For this reason, the coding book that was used in this replication has been included for the use of future researchers who may want to attempt another replication.

Table 3

Coding Categories, Definitions & Examples

Sub-category	Actions items falling into this sub-category with definition	Example from test data
Triggering event	1. **Clarification**	"Clinical nurse specialist (CNS) CNS or nurse practitioners in the acute care setting may be a future trend in nursing in Canada however it is not in the debate topic".
	2. **Restating**	"J said that all staff should be RNs".
Exploration	3. **Agreement –** to concur in (as an opinion) <*agrees* that he is right> plain yes or I agree without substantiation.	"I agree with J".
	4. **Information sharing-** stating a fact, a policy or a rule. - giving information from a reputable source, literature, association website etc.	"Their conclusion was a higher proportion of hours of nursing care provided by RN's and a greater number of hours of care by RN's per day are associated with better care for hospitalized patients"
	5. **Divergence –** to differ in opinion <he *disagreed* with me on every topic> divergent opinion on any point presented by another	"There were a few fallacies in L's remarks. The first one being "enjoying the modern conveniences of PC, I/V pump, and clinical pathways the acuity of the patient is on the rise." Who says nurses are "enjoying" any of these conveniences and what relationship does if have to patient acuity?"
	6. **Leap to Conclusion** – No relationship to previous discussion, not logical	"In fact, the whole team approach through primary care is where we are headed in future health care reform".
	7. **Personal Narration –** story, relating an incident, describing practice in "their" institution	"I remember when I was doing my diploma and degree in nursing, there was no course related to leadership".
	8. **Opinion –** belief or judgment, Personal view, attitude based on grounds insufficient to conclude factual	"I don't think it is ethical to aggressively recruit nurses from underdeveloped countries but my opinion won't matter to the powerful corporations exploiting other countries by scooping their best and brightest".
Integration	9. **Building on** – augmenting a point made by self earlier, or by another	"I was surprised that neither side talked about the use of clinical nurse specialists or nurse practitioners but perhaps it was because these could be grouped into the RN category. But there is an argument that an all RN staffing

		also needs the services of these other specialists in nursing".
	10. **Creating Solution** – novel conclusion	"I'm sorry that I have so many negative comments but the topic alone is one that I have struggled with being in nursing admin. and due to the shortage, was forced to hire other health care workers to assist the RN's".
	11. **Justified hypothesis -** a tentative assumption made in order to draw out and test its logical consequence to prove or show to be just, right, or reasonable -the necessary consequence of two or more propositions taken as premises; coming to conclusion predicted by ongoing discussion but supporting with relevant reason. "therefore"	" The amount of money available for staffing leads good managers to consider a staff mix solution."
	12. **Supported divergence** – disagree "because" - disagree with proof or cause	" I disagree with staff mix <u>because</u> O'Brien-Pallais has proven the increase in patient mortality with decreased percentage of RNs"
	13. **Supported agreement** - to hold up or serve as a foundation or prop for ; agree "because" -agree with proof or cause	I am biased towards L's arguments <u>because</u> implementing the full LPN scope of practice is well under way in specialty areas in my region and it is working well.
Resolution	14. **Wrap-up** – concluding, summarizing	"In conclusion, the debate has demonstrated that the majority of this class supports the affirmative position."
	15. **Thought experiment** – Questioning in a "what if?" fashion or "What do you think about?" – might overlap with Teaching Presence if done by teacher or facilitator to encourage discussion	"What if we put all of the managers in a room for a week and told them not to come out until they balanced the budget?"
	16. **Apply, test, defend** – any one of three but not retrospective narrative. Must be an application of new thought initiated by the discussion present	"I went back to work and surveyed my staff on their position on this issue. They all agreed that …."

Process

There are several ways to divide the data into the segments needed for coding. Some researchers code the entire posting into a category based upon the prevalent action. Others code each sentence. For these data, the RAs were instructed to code every segment of a posting that differed from what preceded it on the basis of the action categories. This could be called coding by "unit of meaning" or "thematic unit" (Rourke et al., 2001). There could be many themes in one posted comment. For example, if a conference participant provided one or 25 different pieces of "factual information" it was one coding, until they presented a new action, such as "drawing a conclusion". Any part of the discussion that did not fall into any of these action item categories was to be ignored. That would include social comments such as greetings and teaching comments such as instructions about responding in the debate.

Coders had been issued copies of Atlas.ti and were trained in its use. Atlas.ti is a software package, developed in Germany. The distributor describes the functions as textual and conceptual. The textual function allows researchers to segment data by categories, essentially

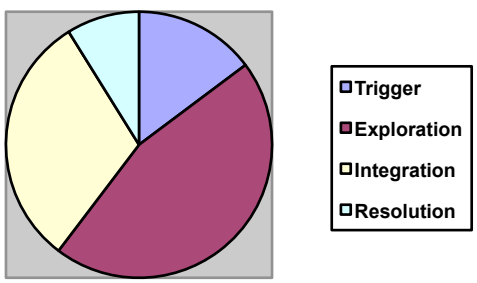
breaking down the data. The conceptual function allows the rebuilding of the data into themes and theories. Detail is provided at www.atlas.ti.com. In practice the software places the entire printed debate down the left hand side of the screen. The PI inserts the action codes into a drop down menu on the screen. Coders high-light segments of the data and assign a given code with a mouse click. The data remain intact but at any time you can look at all of the examples in any action item category. It also automatically counts the number of segments posted into each action item category. It is quite intuitive and the RAs were able to code and then e-mail the final segmented coding to each other and the PI and open it in each individual computer. The RAs did not report any difficulty coding with this software.

After the first round of coding, comparison and discussion took place, which although it did not affect the coding being examined, it might improve the reliability in the second round. One student had coded many more segments than the other, from the same data. For example, if a student debater wrote a long monologue about some position, one student coded it as one clarification and three factual statements while the other found two clarification and six factual statements. It was not easy to determine where this was occurring but the fact that the two students were coding similar percentages of each action item overall was reassuring. This replication continued, with teleconference discussion of interpretations of categories being used as a method to gain corresponding views of how to interpret the meaning of each posting.

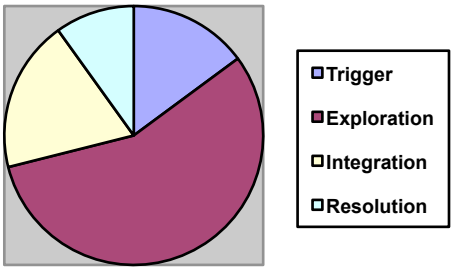
As well, while some action items were easy to identify, i.e. personal narration, others were more difficult, i.e. thought experiment. It is noted in the Computer Human Interaction (CHI) literature that lower levels of cognition are easily identified while higher ones are too complex for simple coding (Sasse, 1997) As our discussion of the interpretation of these action items unfolded, it became clear that the PI could force an interpretation onto the RAs, for each example discussed, to strive for higher inter-rater reliability, but this assumed that the PIs definition was the ‘right’ one. The subjectivity of the higher cognitive levels of communication became an issue which will be discussed in the conclusion. When coding was completed, the PI collapsed the many categories of actions into the four main sub-categories of cognitive presence.

Results

The overall proportional results of the first round of coding by the two RAs are displayed as follows:



Coder #1/Data Set #1



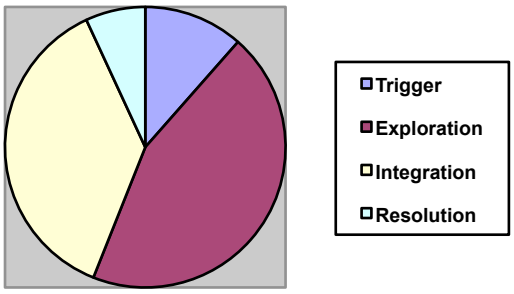
Coder #2/Data Set #1

Figure 1. Pie chart depicting coding data set #1

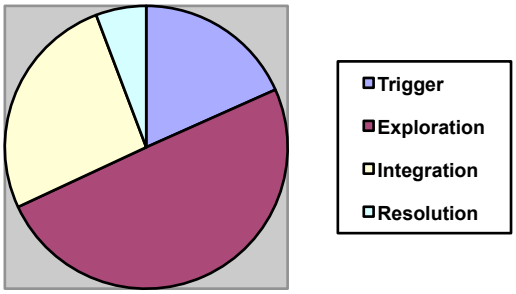
Table 4
Coder Comparisons of Data Set #1

Sub-categories	Coder#1 (items-%)	Coder #2 (items-%)	Inter-rater %
Trigger event	35-14.7	54-14.9	98.6
Exploration	108-45.6	204-56.2	81.13
Integration	73-30.8	69-19	61.68
Resolution	21-8.9	36-9.9	89.9
TOTAL	237	363	Avg. 82.34%

Following the second round of coding, using a new set of data, the coding was again compared using percentages and pie charts for visual effect.



Coder #1/Data Set #2



Coder #2/Data Set #2

Figure 2. Pie Chart depicting coding data set #2

Table 5
Coder Comparisons of Data Set #2

Sub-categories	Coder#1 (items-%)	Coder#2(items-%)	Inter-rater%
Trigger ever	40-11.5	70-18.3	62.8
Exploration	155-44.5	190.49.7	89.5
Integration	129-37	100-26.2	70.8
Resolution	24-7	22-5.8	82.85
TOTAL	348	382	Avg. 76.48%

The individual coders' results were compared as percentages across data sets to examine for patterns in coding, overall. For example, Coder #1, Trigger Event, 78.23% is an average of the percentages of Trigger Events coded in the 1st and 2nd data sets. This is not test-retest as the data sets were different. No pattern emerges.

Table 6
Data Set Coding Correlation-per coder

Sub-categories	Coder#1 (% correlation)	Coder #2 (% correlation)
Trigger event	78.23	81.4
Exploration	97.6	88.4
Integration	83.2	72.5
Resolution	78.7	58.6
AVERAGE	84.4%	75.2%

Discussion

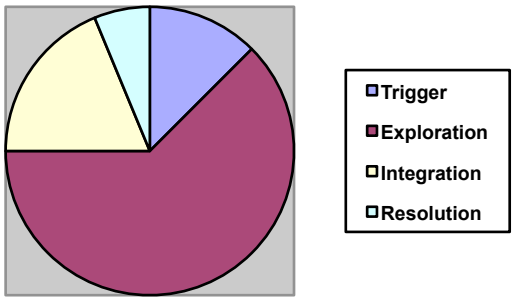
The findings of this replication are that in both sets of data, both coders found that the largest percent of coding represented exploration and the smallest percent represented resolution. Both coders found less exploration and more integration in the second data set than the first. This observation only assumes relevance when compared to finings in other studies.

Garrison et al. (2001) found that 8% of the responses coded in their study, were in the first category - triggering events. Exploration responses composed 42 %, integration 13 %, and resolution 4%. This consisted of 67% of the comments. Thirty-three percent were labelled "other". In the original study "other" represented the categories of social presence and teaching presence, shown in the Community of Inquiry Model. In this study coders were asked to ignore any parts of the discussion which were not cognitive presence action items. For this reason, the "other" category was removed from the Garrison et al. (2001) data depiction and the cognitive presence categories expanded as a proportion of 100%, to make the pie charts comparisons relevant.

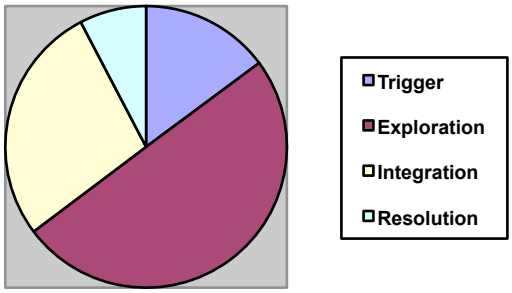
As the original researchers themselves commented, it is interesting that so few messages were coded in what are assumed to be the higher levels of cognition. The instructions to the original coders, who were coding by total posting, were that if it was not clear which sub-category was reflected to "code down" and to "code up" if there was clear evidence of more than one sub-category in the posting (2001, p.11). Inter-rater reliability was tested after three training sessions, with two methods and was not as high as the authors would have liked. They felt that their findings were "groundbreaking" and "rich in analytical value" and represented "a general representation of the frequency of each category", thereby negating the need for higher inter-rater reliability as validation (2001, p.12).

Meyer (2003) published the findings of a study in which she attempted to explain the advantages and disadvantages of traditional in-class vs. on-line education. For part of this research she used what she called the Garrison model (the model used in this study) to analyze on-line threaded discussions for higher order thinking skills. Her findings are depicted in a pie chart below. That study led her to experiment with a selection of coding models for conference postings; the Garrison model then being compared to three other models. Meyer, also includes statistics of percentages of each level of cognitive presence from her studies. Meyer's segment of coding was by entire posting. "If a posting, many of which were quite lengthy, could be

categorized at multiple levels, the level or category most consistent with the entire posting, was used" (2003, p. 107). As all of her coding was done by the researcher herself, there was no attempt at inter-rater reliability. The data from the four studies are very interesting.

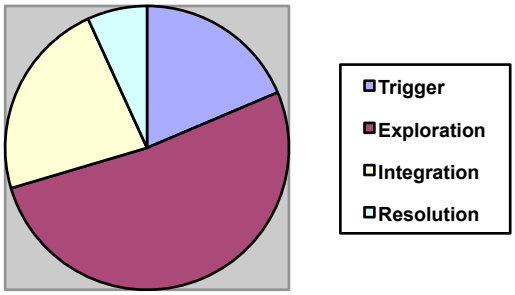


Garrison et al data (2000)

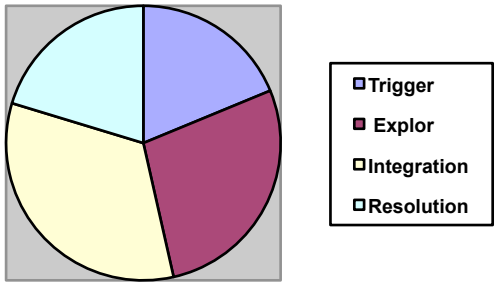


Park data (2003)

Aggregate of all data



Meyer's data (2003)



Meyer's data (2004)

Figure 3. Comparison of fours sets of data

All but Meyer's 2004 data demonstrate a predominance (50% or more) of postings in the exploration category. Meyer explains that the students in her 2004 study were doctoral students whose interaction was in response to "trigger events" that lead to integration and resolution. All of the smallest percentages fall to resolution, the assumed highest cognitive level.

Inter-rater Reliability

Once both sets of data from the two debates had been coded, the plan had been to follow the methods of the original Garrison et al. (2001) study to determine inter-rater reliability. They had compared raters using Holsti's coefficient of reliability and Cohen's kappa. These formulae both require coded segments to be labelled correct or incorrect and the comparisons of numbers of correct and incorrect responses by the coders. The figures recorded by Atlas.ti were total numbers which could be used as percentages but there was no way to compare correct and

incorrect coding from this data. In hindsight it may have been possible to test internal consistency by extracting a small identical portion of data from each coder's work and comparing it to pre-determined responses of correct and incorrect, but this was not done. In the Garrison et al. (2001) study the data consisted of 12 postings made by 4 students. The analyzed debate data contained over 600 coded segments, posted by 12 students. After three training sessions, Garrison et al. (2001) calculated a k of .74. Being unable to calculate those coefficients, percentage agreement was substituted, knowing that it is a less precise measure because it does not factor in the probability of chance. According to Trochim (2002, p.1) "if your measurement consists of categories -- the raters are checking off which category each observation falls in -- you can calculate the percent of agreement between the raters. For instance, let's say you had 100 observations that were being rated by two raters. For each observation, the rater could check one of three categories. Imagine that on 86 of the 100 observations the raters checked the same category. In this case, the percent of agreement would be 86%". In this study inter-rater percentage, per category, ranged from 62.8-98.6. The lowest agreement sub-category was "integration", highlighting the need for explicit descriptors for this category.

Rourke and Anderson (2004) describe the issues that they had encountered when using content analysis coding on computer generated conferencing data, highlights the internal issues relating to this study. These questions underline the need for further study in the field. Does the coding protocol adequately represent a performance domain (i.e. cognitive presence)? Are all performance domains represented in the coding? Are researchers using the coding protocols in the same way? Can a course be designed with "trigger questions" that lead to higher levels of cognitive presence in the CMC? Does cognitive presence increase with educational level? Does researcher bias influence the coding decisions?

What We Learned in the Process

Rourke and Anderson (2004) detail the subjectivity of the construct development in this type of coding and they conclude that there may be multiple plausible interpretations of this type of data or none at all. What had been conceptualized as quantitative data in a coding table was, in this interpretation, very subjective.

From the very beginning of this project, both of the RAs and the PI had been excited about what we were learning about student conferencing. Was that not the point after all? Court (2004) concluded that looking beyond the quantitative content analysis to the qualitative meaning of the exercise is valuable.

The RAs, having taken courses in this program, immediately started to examine their own postings, and those of their classmates, in a different light. They said that they tried to use what they assumed to be higher levels of cognition, as found in the coding tool. As an instructor in such a course, I now found it much easier to identify the types of content that individual students presented. Although I did not use this to evaluate student participation, I was able to indicate to students that, for example, all of their postings were exploratory and perhaps they could attempt to develop and test a concept that they were learning. I went back to Court's (2004) comments that "we want to learn about effective programs and teaching methods to help students learn" (p. 1) but in doing this, I realize that some of the discussion topics posted in my own course did not lead to integration and resolution. I realized that the "right" kind of input, i.e. trigger events, could stimulate higher percentages of a more advanced level of the cognitive tool. Meyer (2004) supports this contention. She tested four different "frames", or models, for assessing student conferencing, one of them being the tool used in this study, and concluded that "each frame has

value", yet "each frame focuses attention on some particular aspect or quality of the student" (p.112). She did, however, conclude that, "the type of trigger question may generate the level of response from the students"(p.112).

Research Findings

In this study, the Cognitive Presence Tool developed by Garrison, Anderson and Archer (2001) was looked at through "new" eyes. The sub-categories were interpreted by the PI and two coders using descriptions from the original study, dictionary definitions, discussion and consensual decisions. This interpretation of the tool was used to code the data of this study using the same content analysis software as used in the original study.

The process of data collection was easily replicated. The content analysis was not. It became evident to this team very early on that the definitions of categories of cognitive presence were subjective, although no one had difficulty coding using their perceived interpretations. We had no way to evaluate if each of us was interpreting the data entirely the same way and absolutely no assurance that we were interpreting the categories in the same way as the original researchers. We also were unable to complete the same inter-rater reliability testing as described in the original study. The percentage comparisons presented in this report do indicate a high relationship between the coding of the two coders and for each of them between the two sets of data. As well, the percentages of each sub-category of cognitive presence in Garrison et al (2001) Meyer's first study (2003), and this data, are remarkably similar.

Limitations

1. In both the original study and in the replication the sample size was small and the numbers of data elements per course varied.
2. In both the original study and in the replication the determination of coding category allowed for subjective interpretation.
3. The definition of each coding unit can have a subjective nature.
4. The "Trigger Event" can affect the type of response obtained from students.
5. The same types of inter-rater reliability were not attempted so comparisons of reliability are not possible.

Conclusions

Can the Cognitive Presence Tool, developed by Garrison, Anderson and Archer (2001a, 2000 b; 2001), be successfully applied to different data by a different research team? Yes, the physical mechanics of the process were replicable.

What type of replication was achieved? A strict replication was not achieved. Although there were similarities in the subject characteristics, research histories of subjects, historical context and physical setting of research, we were unable to show that the interpretations of the data by coders in this study bears any relationship to the interpretations of the codes in the original study, because of the subjective nature of the higher cognitive level concepts. According to Hendrick's (1990) definitions, a conceptual repetition, "an attempt to convey the same crucial structure of information ... but by a radically different transformation of the variables" (p. 45) or a partial replication, "some change to part of the procedural variables" (p. 45) was achieved. We replicated the process but not all of the analysis. He claims that conceptual and partial replications have some logical status, but not that of a strict replication. The status achieved by this replication attempt is in the identification of issues that need analysis and research to improve the development of cognitive assessment tools.

Garrison, Anderson and Archer (2001) concluded that, "this could be a valuable tool for teachers to assess and confirm the nature of the discourse appropriate for the desired learning outcomes". They also stated that this could "serve as a framework for future research into a much needed quest to better understand the cognitive nature of the teaching and learning transaction and asynchronous text-based conferencing environment"(p. 14). In an earlier publication, this group, along with a graduate student, Liam Rourke (2001), commented on the lack of replication in content analysis of communication studies. They stated that "unfortunately, few researchers appear to be interested in conducting their studies with existing instruments." (p.16). They also reported that Henri's (1991) protocol for coding had been reused but had been modified each time. (p.12) This study then, is an attempt to replicate the Garrison, Anderson and Archer (2001) study, using a similar protocol.

The disappointment experienced by the researchers in this replication study, in relation to the issue of inter-rater reliability and the concept of never coming to complete agreement on definitions for the higher level cognitive activity, was tempered by reading that, in the emerging field of Human Computer Interaction (HCI), there is a view that "the temptation to formalise phenomena before they are sufficiently well understood, and abstracting complex phenomena (such as human reasoning) into simplistic, mechanical models, leads to squeezing square pegs into round holes rather than rethinking the system and developing a better approach" (Sasse, 1997, p.1) Continuing research on the Garrison et al. (2001) model must focus on the description and meaning of the higher meta-cognitive practices that are occurring in CMC but are difficult to identify. Anderson (2003) himself has moved away from this model in his quest to conceptualize interaction as a component in learning, by developing new theories and models.

The activity of replication itself is very instructional and expands one's ability to communicate with researchers in the field under study about their actions and beliefs. It is recommended that anyone attempting to expand a conceptual field of research seriously consider replicating some of the work of those who have published in the same field previously. If nothing else, it will allow researchers to determine if they are on the same "wave-length" as the pioneers. Through this exercise, the goals of the Centre pilot were achieved. Access to archived course discussion database as research data was accomplished. Computer software for content analysis was successfully understood and used by PI and coders and graduate students, at a geographic distance from the PI, were hired, trained and supervised, receiving and sending software and data electronically.

Future Research

The research questions relating to replication in this instance are complete, but the concept of replication is still as important as stated in the beginning (Park, 2004). Educational innovation will increase as researchers and teachers are able to consistently demonstrate improved outcomes of pedagogically sound technology use. The data provided by online discussion based courses leads in many directions. The data can me mined from the perspectives of the student, the teacher or the content, as demonstrated by the Community of Inquiry Model. As well, pre and post testing the introduction of a new teaching technique is greatly facilitated with this data source.

The data also lends itself to informing qualitative research questions such as themes of interaction and relationship (Schrire, 2006).

References

Anderson, T. (2003, October). Getting the mix right again: An updated and theoretical rationale for interaction. *International Review of Research in Open and Distance Learning*. Retrieved April 06 2009, from http://www.irrodl.org/index.php/irrodl/article/view/149/708

Court, D. (2004). The quest for meaning in educational research. *Academic Exchange Quarterly, 8*(3), 1-6.

Garrison, R. D., Anderson, T., & Archer, W. (2000a). Critical inquiry in a text-based environment: Computer conferencing in higher education. *Internet and Higher Education*, *11*(2), 1-14. Retrieved April 06 2009, from http://communitiesofinquiry.com/files/Critical_Inquiry_model.pdf

Garrison, R. D., Anderson, T., & Archer, W. (2000b). Critical thinking, cognitive presence, and computer conferencing in distance education. Unpublished manuscript found April 6, 2009 at http://communitiesofinquiry.com/files/CogPres_Final.pdf

Garrison, R. D., Anderson, T., & Archer, W. (2001). Critical thinking and computer conferencing: A model and tool to assess cognitive presence. *American Journal of Distance Education, 15*(1), 1-18.

Hendrick, C. (1990). Replications, strict replications, and conceptual replications: Are they important? In J. W. Neuliep (Ed.), *Handbook of replication research in the behavioral social sciences* (pp 51-70). Madeira, CA: Selected Press.

Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussion. *The American Journal of Distance Education, 17*(1), 25-43.

Meyer, K. A. (2003). Face-to-face versus threaded discussions: the role of time and higher order thinking. *Journal of Asynchronous Learning Networks, 7*(3), 55-65.

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

Park, C. L. (2004). What is the value of replicating the studies of others? *Research Evaluation, 13*(3), 189-195.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education*, *12*.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development, 52*(1), 5-18.

Sasse, M. (1997). *Eliciting and describing users' models of computer systems*. Unpublished doctoral thesis, University of Birmingham, England.

Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education*, *46*(1), 49-70.

Trochim, W. M. K. (2002). *Research methods knowledge base.* Retrieved April 6, 2009, from http://www.socialresearchmethods.net/kb/reltypes.htm

Notes.
The contributions of the two RAs, Leah Keats and Corinne Katazinski, who were graduate students in the MHS program at the time of the study, are greatly appreciated.