JOURNAL OF DISTANCE EDUCATION
REVUE DE L'ÉDUCATION À DISTANCE
SPRING/PRINTEMPS 2005
VOL. 20, No 2, 1-20

ARTICLES

Evaluating Critical Thinking Skills: Two Conceptualizations

Cheryl L. McLean

Abstract

In this study two complementary models, which were based on the strengths of existing models, were developed to analyze students' critical thinking skills. One model was used to categorize the types of critical thinking displayed by students; the other was used to evaluate the quality of the critical thinking. The models were refined and tested for usefulness through an analysis of students' messages posted in two sequential two-week computer conferences. The study also examined the level of critical thinking displayed compared with that shown by other studies and highlighted critical thinking skills that were not frequently demonstrated by participants.

Résumé

Dans cette étude, deux modèles complémentaires, basés sur les points forts de modèles existants, ont été développés pour analyser la capacité de pensée critique des étudiants. Un modèle a été utilisé pour catégoriser les types de pensée critique démontrés par les étudiants, alors que l'autre a été utilisé pour évaluer la qualité de la pensée critique. Les modèles ont été raffinés et testés en fonction de leur utilité grâce à une analyse des messages des étudiants placés dans deux forums numériques séquentiels de deux semaines. L'étude a aussi examiné le niveau de pensée critique démontré en comparaison avec celui indiqué dans d'autres études et a mis en lumière des habiletés de la pensée critique peu fréquentes chez les participants.

Halpern (1998) points out that "there is virtually no disagreement over the need to help college students improve how they think" (p. 450). The literature on critical thinking, however, suggests that many college students are not performing well on critical thinking tasks (King & Kitchener, 1994; Paul, 1993). Adding to the difficulty is the fact that there is no generally accepted definition of critical thinking; nor is there a generally accepted model to evaluate critical thinking.

Several well-known authors have developed definitions of critical thinking. Ennis (1987) bases his definition on five key ideas: "practical, reflective, reasonable, belief, and action" (p. 10), resulting in a working

definition of: "Critical thinking is reasonable reflective thinking that is focused on deciding what to believe or do" (p. 10).

Brookfield (1987) focuses on assumptions. His definition "involves calling into question the assumptions underlying our customary, habitual ways of thinking and acting and then being ready to think and act differently on the basis of this critical questioning" (p. 1). Both definitions have elements of both *product* and *process*. The *product* is the decision made about thinking or acting through the *process* of reflection or questioning. Although Brookfield points out that critical thinking is a process and not an outcome, this does not negate the product portion of his definition, which is thinking or acting differently. Brookfield's definition, with its emphasis on assumptions, appears to be a little narrower than Ennis'.

Paul's (1993) definition is focused on the process. He defines critical thinking as

> a unique kind of purposeful thinking in which the thinker systematically and habitually imposes criteria and intellectual standards on the thinking, taking charge of the construction of thinking, guiding the construction of the thinking according to the standards, [and] assessing the effectiveness of the thinking according to the purpose, the criteria, and the standards. (p. 21)

His description of the process is more detailed than either Ennis' or Brookfield's, but his definition lacks an outcome or product.

Selecting a definition of critical thinking even from the few listed above is difficult. Some, such as Ennis (1987), focus on both product and process. For others, such as Paul (1993) and Brookfield (1987), the primary focus is on process. All the process definitions highlight specific critical thinking skills, but do not cover all aspects of critical thinking. Lipman (1991) points out that the current definitions are too vague and fail to note the characteristics of critical thinking. However, it seems unreasonable to expect a single definition to cover all the competences that might be displayed by critical thinkers. Perhaps the definition is not the place to delineate the specific skills. A model of critical thinking skills seems to be the more appropriate venue to list and define critical thinking skills and competences.

Models of Critical Thinking

The primary purposes of a model are to provide an accurate view of the phenomena under study and to facilitate communication about those phenomena. Ennis (1987) has designed a taxonomy of critical thinking dispositions and abilities. He lists 12 abilities that represent four basic areas of critical thinking: "clarity, basis, inference, and interaction" (p. 16). Ennis divides clarity into two groups: elementary and advanced, and

includes focusing on a question, analyzing arguments, and asking and answering questions of clarification and/or challenge in the elementary category. Advanced clarification includes defining terms and judging definitions and identifying assumptions. Basis, Ennis' second basic area of critical thinking, refers to the abilities to support one's inferences and to assess evidence. He includes judging the credibility of the source and observing and judging observation reports. The third area, inference, includes deducing and judging deductions, inducing and judging inductions, and making value judgments. The final area, interaction, focuses on interacting with others and deciding on an action. The latter includes: "a) define the problem, b) select criteria to judge possible solutions, c) formulate alternative solutions, d) tentatively decide what to do, e) review, taking into account the total situation, and decide, f) monitor the implementation" (p. 15).

Ennis' (1987) model provides a detailed list of critical thinking abilities. Criteria by which to judge the quality of these abilities, however, are missing. For example, one ability is "inferring explanatory conclusions … [by interpreting] authors' intended meanings" (p. 13). It is possible to make such an inference and still not display good critical thinking if the inference is at a surface level or if the interpretation is inaccurate. Although it appears that this model was designed more to teach critical thinking than to evaluate it, the model may serve as a strong base from which to develop a model suitable for categorizing the types of critical thinking skills demonstrated by students. Another means of evaluating the quality of the skills would have to be found.

Brookfield (1987) sees four components to critical thinking: identifying and challenging assumptions, challenging/recognizing the importance of context, imagining and exploring alternatives, and engaging in reflective skepticism. He points out that critical thinkers are wary of claims of universal truth. Their understanding of assumptions, context, and alternatives makes them reflectively skeptical of ultimate explanations.

Brookfield (1987) identifies five phases of critical thinking: the trigger event, appraisal, exploration, developing alternative perspectives, and integration. These have elements in common with Ennis' "deciding on an action" ability. In the first phase, the trigger event, unexpected events occur that result in a sense of "inner discomfort and complexity" (p. 26).

The second stage, appraisal, is similar to Ennis' "defining the problem and selecting criteria to judge solutions," but seems to have greater depth. In this stage the thinker appraises the situation. This appraisal may include self-scrutiny, minimization and denial, a focus on the nature of the problem, identification and clarification of the problem, and a search for others with a similar problem.

In the third stage, exploration, the thinker begins to look for and test new ways of explaining or dealing with the situation. This stage is similar to Ennis' "formulating alternative solutions."

Developing alternative perspectives is the fourth stage in which the thinker selects the solution to the problem that seems to be the most appropriate and that will fit in best with his or her ways of thinking and living. This matches Ennis' "tentatively deciding what to do."

Integration is the final stage, in which the solution selected as the most appropriate in the previous stage is integrated into the thinker's life. The solution may involve a change or it may involve a renewed commitment to an already existing stance. Although this phase seems to have more depth than Ennis' "reviewing the solution and monitoring the implementation," both focus on the results of the decision.

Brookfield's (1987) components have a narrower focus than Ennis' model, emphasizing assumptions, context, and alternatives, resulting in more depth in these areas. An important concept that does not appear in Ennis' model is the idea of reflective skepticism. Brookfield treats it as an advanced critical thinking skill that would seem to be an important factor in Ennis' deciding what to believe or do. Brookfield's model appears to have the quality of thinking integrated into the components. For example, critical thinkers would reject inappropriate assumptions and seek new ones.

Integrating reflective skepticism and adding some of Brookfield's depth to the Ennis model may result in a more useful model.

King and Kitchener's (1994) seven-stage reflective judgment model "describes a developmental progression that occurs between childhood and adulthood in the ways that people understand the process of knowing and in the corresponding ways that they justify their beliefs about ill-structured problems" (p. 13). An important idea here is the concept of an ill-structured problem. King and Kitchener are quite clear that unless the problem is ill-structured and does not have an obvious solution, critical thinking or reflective judgment is not required. The seven stages are grouped into three primary categories: pre-reflective thinking, stages one, two, and three; quasi-reflective thinking, stages four and five; and reflective thinking, stages six and seven.

King and Kitchener (1994) have found that first-year college students typically score in stages three and four. Those in stage three view knowledge as certain or only temporarily unknown. In their concept of justification, the authorities' views are seen as paramount where knowledge is known. Personal opinion is important in knowledge that is temporarily unknown. In stage four students begin to use evidence to support their judgments, and they are beginning to recognize that knowledge is uncertain. They assume that there could be many possible answers to an ill-

structured problem and that an appropriate solution is a matter of personal opinion only. In stage five, students understand that an individual's point of view or perspective may influence that individual's interpretation of evidence, resulting in knowledge that is contextual and subjective. Beliefs are justified in a particular context.

Stages six and seven represent reflective thinking. In stage six, knowledge is seen as constructed individually, based on information from a variety of sources. Beliefs are justified by comparing evidence and opinion from varying perspectives on an issue or across varying contexts by constructing solutions that are evaluated by criteria such as "the weight of the evidence, the utility of the solution, or the pragmatic need for action" (p. 15).

Stage seven represents the highest level of reflective thinking and was rarely seen in college students. In this stage, knowledge is seen as the outcome of a process of reasonable inquiry. The solutions to ill-structured problems are evaluated according to current evidence.

> Beliefs are justified probabilistically on the basis of a variety of interpretative considerations, such as the weight of the evidence, the explanatory value of the interpretations, the risk of erroneous conclusions, consequences of alternative judgments, and the interrelationships of these factors. (p. 16)

These stages are meant to cover a lifetime of critical thinking, and thus the entire model is not likely to be useful in evaluating critical thinking over a single course. It seems worthwhile, however, to look for the skills described in stage seven as part of a high level of critical thinking.

Henri and Rigault's (1996) model was developed specifically to evaluate interaction in computer conferences in distance education. They used an analysis grid to analyze the speech segments in an interaction among learners or between learners and the instructor. Each speech segment was analyzed in terms of content, function, characteristics, and author. The section that dealt with the cognitive function may be useful for evaluating critical thinking. Henri and Rigault divided the cognitive function into two levels: *surface* and *in-depth*. They further divided the surface level into two components: repetition of what someone else has said and subjective value judgment and the in-depth category into three components: clarification, interpretation, and value judgment. There are some clear parallels between this model and Ennis' (1987) model. In both cases *clarification* is a primary category. Whereas Ennis focuses on types of activities that might be seen in more formal critical thinking, Henri and Rigault include activities that might appear in less formal situations such as delving deeper, broadening the scope, and reformulation of the problem through personal example or translation. Henri and Rigault's second

primary category, interpretation, covers many of the activities that Ennis lists in his *inference* category. Henri and Rigault have placed value judgments in their own primary category, unlike Ennis who has included them with the *inference* category.

The strength of this model is the indicators of each category, which have been developed through observation of interactions in computer conferencing. The categories, however, do not have the level of detail found in Ennis' model. The evaluation of the quality of the thinking seems to be in the in-depth and surface categorizations.

According to Paul and Elder (n.d.), comprehensive critical thinking is responsive to and guided by intellectual standards. The standards include clarity, accuracy, precision, relevance, depth, logic, and breadth. These standards clearly address the quality of thinking, a component missing in the other models.

Two models may be required to gain a clear picture of the critical thinking displayed by students: one to categorize the kinds of critical thinking displayed and one to evaluate the quality of the critical thinking displayed. Rather than selecting one model for the categories of critical thinking, it may be possible to draw on the strengths of several models and develop an integrated model. Ennis' abilities section of his detailed curriculum model is a good starting point. Adding Henri and Rigault's making judgments category to Ennis' major categories would allow for a stronger focus on the final step of the critical thinking process, in keeping with the model of King and Kitchener. Brookfield's reflective skepticism is an important component that should be included, as are the high level critical thinking skills in King and Kitchener's stages six and seven.

Elder's and Paul's (n.d.) universal intellectual standards (clarity, accuracy, precision, logic, relevance, depth, and breadth) provide a means of evaluating the quality of thinking displayed in each of the categories of critical thinking demonstrated by students.

Critical Thinking in Computer Conferencing

The importance of interaction and discussion among learners in promoting critical thinking skills is generally recognized among both theorists and practitioners (Brookfield, 1987; Ennis, 1987; Henri, 1995; McPeck, 1990; Newman, Webb, & Cochrane, 1995). As Garrison (1992) notes, "Meaning developed in isolation does not meet the criteria of critical or reflective learning. Critical discourse is essential for worthwhile and valid knowledge" (p. 139).

One of the most frequent criticisms leveled at postsecondary distance education in the past, however, has been the isolation of the learner and the lack of opportunities to build and test knowledge structures and develop critical thinking skills through interaction with other learners and

the instructor. Computer conferencing provides opportunities for this interaction and, therefore, opportunities for developing critical thinking skills. As Hara, Bonk, and Angeli (2000) point out, "there is a pressing need to consider the dynamics of the online discussion and how it may facilitate students' cognitive and metacognitive development" (pp. 115-116).

Studies on critical thinking in computer conferences are emerging. Burt, Grady, and McMann (1994) examined the interaction among graduate students in inter-university computer-mediated conferences. For content analysis they used Henri's (1992) model, which includes elementary clarification, in-depth clarification, inference, judgment, and strategies. The analysis showed a high level of clarification statements. Students reached a level of information sharing but not knowledge construction.

Bullen's (1998) study examined participation and critical thinking in a university-level computer conference, using a modified form of Ennis' critical thinking model. His primary categories were clarification, assessing evidence, making and judging inferences, and using appropriate strategies and tactics. He used positive and negative indicators to judge the quality of each skill. Bullen found low to moderate participation levels, and although all students demonstrated critical thinking at some level, none was doing so at the highest level on a consistent basis. He pointed out that the relatively passive role of the moderator may have contributed to the low level of critical thinking and participation.

Gunawardena, Lowe, and Anderson (1997) examined a debate in an online computer conference. Participants were distance education professionals debating the "role and importance of 'interaction' in effective distance education" (p. 401). The primary purpose of this study was to construct an interaction analysis/content analysis model to examine the "negotiation of meaning and co-construction of knowledge in collaborative learning environments facilitated by computer conferencing" (p. 397). The resulting model consists of five major phases, including

> sharing/comparing of information.… the discovery and exploration of dissonance or inconsistency among ideas, concepts or statements … negotiation of meaning/co-construction of knowledge … testing and modification of proposed synthesis or co-construction … agreement statement(s)/applications of newly constructed meaning. (p. 414)

This model focused strongly on the group process and would be less useful in evaluating individuals' thinking skills. Most of the postings indicated a fairly high level of thinking. The participants in this study may not be typical as they were professionals who would be expected to have this level of skill.

Kanuka and Anderson (1998) examined the level of thinking of workplace training managers discussing learning technologies on a computer conference using the same model as Guanwardena et al. (1997). They found little evidence of higher levels of thinking. Gabriel, MacDonald, Farres, and Stodel (2000) also looked at the applicability of the Gunawardena et al. model, examining the online communications of students in MBA graduate programs. Most postings were at the lower levels of comparing and sharing information. They also found that a large majority of the postings were social interaction.

Hara et al. (2000) analyzed a computer conference in a graduate-level course using Henri's (1992) model of content analysis, with some modifications. Their categories included elementary clarification, in-depth clarification, inferencing, judgment, and application of strategies. They found that students showed good critical thinking skills, including inferencing and judgment. In examining the level of information processing, they found that 55% of the student messages were at the in-depth level, and 33% were at the surface level.

McKenzie and Murphy (2000) also used a modified form of Henri's (1992) model in their study of a graduate-level discussion group. About half the messages were either in-depth or elementary clarification. Inference and judgment were much lower at 7% and 19% respectively. In looking at the level of information processing, 67% of the message units were classified as deep processing, and 22% were classified as surface processing.

As can been seen from these studies, a variety of definitions of critical thinking are used in the current research, and no single model of analysis predominates. Therefore, the purpose of this study was to conceptualize a means of analyzing critical thinking that would allow users both to categorize the types of critical thinking displayed and to evaluate the quality of the critical thinking. Two different but complementary models were developed based on the strengths of the models described above. The models were refined and tested for usefulness through an analysis of students' messages posted in two sequential two-week computer conferences.

Method

The participants in the study were university students enrolled in an undergraduate correspondence psychology course in a Canadian university. A letter was sent to the students offering them an opportunity to participate in two computer conferences in lieu of the three written assignments normally expected in this course. Of the 58 students enrolled in the course, 32 students chose to participate in the study.

Two conference questions were developed. Ill-structured questions to which there were no obvious answers were selected to promote as much critical thinking as possible.

Two models to evaluate critical thinking were prepared: one to evaluate the quality of students' critical thinking and one to determine the categories of critical thinking that students displayed. Both models are described in the data analysis section below.

Pre- and postconference participant questionnaires were developed to gather students' demographic information as well as information on students' perceptions of the effectiveness of the computer conferences in developing critical thinking skills.

Procedures

Students who chose to participate were randomly assigned to one of three groups, each of which had its own private conference area.

The introductory conferencing activities (on-line introductions and the preconference questionnaire) lasted 10 days and were designed to ensure that all students could access and use the computer conferencing system. Two sequential computer conferences followed, each lasting two weeks. All three groups received the same questions, one question for each conference. Participants were required to post a minimum of five messages in each conference.

Data Analysis

The participants' messages posted and stored on the conferencing system formed the main data set. The technique used to analyze these data was quantitative content analysis, which Riffe, Lacy, and Fico (1998) describe as "the systematic assignment of communication content to categories according to rules, and the analysis of communication involving those categories using statistical methods" (p. 2). The categories to which the content was assigned were the components of the two critical thinking models.

Quality of Critical Thinking Model

An adapted form of Paul and Elder's (n.d.) list of criteria, including clarity, accuracy, precision, relevance, depth, breadth, and logic, was used as it focuses on specific criteria. A more detailed description of the model used at the beginning of the analysis is shown in Table 1. It has seven criteria. *Support* was added to the list because one of the course learning objectives focused on students being able to support their point of view. *Accuracy* was eventually removed as it was not possible for non-subject-matter specialist coders to determine a definitive score for this criterion.

Table 1
Quality of Critical Thinking Model

Criteria	*Indicators*
Clarity	Is the point the student is trying to make clear? Does it need further elaboration, examples, or illustrations?
Relevance	Is the message focused on the main topic of the conference?
Depth	Does the student's response address the complexities of the question, or is the student working at the surface level? Is the student bringing in a new idea or taking an existing idea into new territory? Is the student advancing the discussion?
Logic	Is the participant demonstrating a logical argument?
Precise	Does the participant elaborate on the main point he or she is making? Has the participant provided the relevant details?
Breadth	Does the participant consider other points of view? Does the participant look at the question in more than one way?
Support	Are the participant's statements supported by references to credible sources or clearly described personal observations?

(Adapted from Paul & Elder, n.d.).

Students' messages were divided into discrete *units of meaning* that encompassed a single topic, normally a paragraph. For each of the seven criteria a point was given if the unit of meaning was a positive example of that criterion. Two interrater reliability tests were conducted: percentage of agreement and Pearson product moment correlation. A score of 80% or better is acceptable for the percentage of agreement test (Riffe et al., 1998). Pearson product moment correlation scores between .70 and .75 are considered acceptable (Banerjee, Capozzoli, McSweeney, & Sinha, 1999; Mitchell & Jolley, 2001). For conference 1, the percentage of agreement was 89%, and the Pearson product moment correlation was .74. For conference 2, the percentage of agreement was 92%, and the Pearson product moment correlation was .75.

A mean critical thinking score out of seven was calculated for each student for each conference. For example, a student with a mean score of 3.03 was demonstrating on average approximately three of the seven criteria in each analysis unit.

Categories of Critical Thinking Model

A categories of critical thinking model was developed based on the models of Ennis (1987), Brookfield (1987), Henri and Rigault (1996), and King and Kitchener (1994). A portion of the model is shown in Table 2. Its four primary categories are: (a) clarification of the thesis, problem, or

Table 2
Categories of Critical Thinking Model: Levels One and Two

Level one	*Level two*
I. Clarification of the thesis, problem, or question	Reformulating/translating Summarizing Questioning Defining terms Identifying and challenging assumptions
II. Making inferences and interpretations	Deductions Inductions Imagining and exploring alternatives Transfer to other situations
III. Supporting inferences and interpretations	Using credible sources Using personal observations
IV. Making value judgments	Evaluating the validity and interpretability of the source or supporting material Evaluating observation reports Considering the adequacy of the solution, inference, or interpretation Reevaluating when new evidence, perspectives, or tools of inquiry become available and change a position when the evidence and reasons are sufficient to do so

(Adapted from Ennis, 1987; Brookfield, 1987; Henri & Rigault, 1996; King & Kitchener, 1994).

question; (b) making inferences and interpretations; (c) supporting inferences and interpretations; and (d) making value judgments.

Each analysis unit was examined for examples of the four primary categories. As one or more examples of a category were found, that category was given a single point for that analysis unit. Interrater reliability tests were conducted using percentage of agreement, resulting in scores of 95%.

Percentage scores were calculated for each participant for each of the four primary categories. For example, a participant receiving a score of 43% for *supporting inferences and interpretations* made those types of statements in 43% of the analysis units.

Results

In conference 1, 32 participants left 217 messages for a total of 58,431 words. In conference 2, 32 participants left 211 messages for a total of 52,111 words.

Table 3
Quality of Critical Thinking Scores by Treatment Group by Conference

			Critical thinking score		
	Group	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
Conference 1	1	2.67	.49	1.86	3.57
	2	3.14	.73	2.13	4.42
	3	2.84	.68	1.91	4.08
Conference 2	1	3.46	.49	2.71	4.33
	2	2.98	.52	2.25	4.18
	3	3.03	.55	2.25	4.29

Quality of Critical Thinking Analysis

Table 3 shows the descriptive statistics for each group for each conference. The individual criteria in the quality of critical thinking model were also examined. The percentage of thought units in which each criterion was given a positive score for each group for each conference were evaluated as shown in Table 4. For example, group 1 received a point for *clarity* in 54.02% of the analysis units. The category most frequently given a point for all groups in both conferences was *relevance*. The category least frequently given a point for all groups in both conferences was *breadth*.

Categories of Critical Thinking Analysis

An analysis was conducted on the categories of critical thinking displayed by the participants. Table 5 shows the descriptive statistics for each category of critical thinking for each group for each conference. The category that appeared in the analysis units most frequently in both conferences for all

Table 4
Positive Score Percentages for the Quality of Critical Thinking Model Criteria by Group by Conference

		Conference 1			*Conference 2*	
Criteria	*Group 1*	*Group 2*	*Group 3*	*Group 1*	*Group 2*	*Group 3*
Clarity	54.02	62.61	59.08	75.61	61.46	63.31
Relevance	98.35	99.47	93.83	100.00	99.24	95.59
Logic	13.75	21.29	24.09	29.10	15.57	17.50
Precise	9.94	19.05	12.22	30.07	14.92	14.51
Depth	82.53	88.88	78.40	94.23	88.88	87.92
Breadth	3.15	04.61	03.54	4.39	01.58	02.49
Support	05.42	18.27	13.01	12.51	16.16	21.31

Table 5
Categories of Critical Thinking Percentage Scores by Group by Conference

	Conference 1						*Conference 2*					
	Group 1		*Group 2*		*Group 3*		*Group 1*		*Group 2*		*Group 3*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Category												
Clarification of the thesis, problem, or question	38.42	7.99	41.58	14.60	36.01	16.31	37.82	13.40	46.03	15.74	40.28	19.96
Making inferences and interpretations	59.14	9.14	77.06	12.38	59.37	17.69	51.84	15.20	75.88	10.74	57.13	21.48
Supporting inferences and interpretations	9.15	7.01	23.02	8.88	13.43	7.16	8.62	7.99	25.02	18.90	18.56	9.49
Making value judgments	7.23	6.97	7.59	5.26	13.71	11.97	7.84	5.56	14.93	11.50	10.10	7.82

Note. The scores in the above table are percentages.

groups was *making inferences and interpretations* followed by *clarification of the thesis, problem, or question*. The category that appeared the least was *making value judgments*.

Postconference Questionnaires

The participants completed a 21-item questionnaire. Table 6 shows the summary results of the multiple-choice questions. The open-ended questions were analyzed using qualitative analysis techniques as described in Miles and Huberman (1994). Several themes were shared by all three groups, including the positive value of the interaction among the participants, seeing multiple perspectives, and expressing their own opinions. All three groups also talked about enjoying learning through computer conferencing.

The reasons given by group 3 for a lower level of satisfaction with conference 2 were examined. Eight of 11 participants rated their satisfaction as moderate, and one rated it as low in conference 2. Of those who rated it moderate, two participants actually rated it higher than the first conference, which they rated low. Excluding these two participants, the reasons given for moderate or low ratings included the difficulty of the topic (4) and not liking the topic (2).

Table 6
Participant Postconference Questionnaires: Multiple Choice Questions Summary

Item	*Rank*	*Group 1*	*Group 2*	*Group 3*
Question 1: What was your level of satisfaction with the first conference	Low	0%	0%	18%
	Moderate	30%	18%	9%
	High	70%	82%	73%
Question 3: What was your level of satisfaction with the second conference?	Low	10%	0%	9%
	Moderate	30%	27%	73%
	High	60%	73%	18%
Question 5: Did your critical thinking skills improve as a result of the conference?	Yes	60%	73%	55%
	No	20%	0%	27%
	Undecided	20%	27%	18%
Question 7: Did the conferences contribute to your learning in this course?	Yes	60%	64%	55%
	No	10%	18%	9%
	Undecided	30%	18%	36%
Question 9: Do you feel that the conferences were a good use of your time?	Yes	70%	91%	55%
	No	20%	0%	27%
	Undecided	10%	9%	18%
Question 14: Would you participate in another computer conference?	Yes	80%	82%	91%
	No	10%	9%	9%
	Undecided	10%	9%	0%
Question 17: Did you encounter any technical difficulties?	Yes	10%	36%	36%
	No	90%	64%	64%
Question 19: Did you enjoy the conferences?	Yes	100%	91%	82%
	No	0%	0%	18%
	Undecided	0%	9%	0%

Discussion

Quality of Critical Thinking Model

As can be seen from the descriptive statistics, the quality of critical thinking scores were in the low to moderate range: 2.89 to 3.14. This finding is consistent with other studies including Bullen (1998), Kanuka and Anderson (1998), Hara et al. (2000), and Rose (2004). Other studies did show higher levels (Gunawardena et al., 1997; McKenzie & Murphy, 2000), but they were at the professional or graduate level. Gibson's (1996) study may provide a clue as to why studies at the undergraduate level are not showing high levels of critical thinking. She suggests that the quality of thinking progresses across time from content-based skills to critical and complex thinking skills and that that we should not apply inappropriately high standards at the beginning of a conference or series of conferences. Perhaps two conferences over a four-week period is simply not enough time

for students to move from low and moderate levels of critical thinking to higher levels.

According to the descriptive statistics in Table 4, the criteria that received the least frequent positive scores for all three groups were *breadth* and *precise*. The percentage of time that students were not demonstrating *breadth* is startling at 95-98%. A lack of *breadth* indicates that individual participants were not looking at the question from a variety of viewpoints, nor were they considering alternative explanations, which is somewhat surprising given the *appreciation for multiple perspectives* theme that appeared in the participants' postconference questionnaires. A lack of *precision* generally indicated that participants were not providing sufficient detail in their answers or that their answers were vague. Although somewhat better than *breadth*, the percentage of analysis units in which this criterion was not demonstrated was also very high at 70-90%. The criterion *support* was one of the three lowest scoring criteria in conference 1 for all three groups and one of the three lowest-scoring criteria in conference 2 for group 1. This low score indicated that participants were not supporting their statements with evidence or were supporting them inaccurately 79-95% of the time. In conference 2, the criterion that had one of the three lowest frequencies for groups 2 and 3 was *logical*. A lack of logic could mean that the argument the participant presented was not logical. It could also mean, however, that an argument was not presented at all. The participant may have presented an opinion with no logical reason given for that opinion. *Logical* was given a negative score in 70-86% of the analysis units.

Relevant and *depth* were the two criteria that most frequently received a positive score. *Relevant* ranged from a low percentage of 94 in group 3 for conference 1 to a high of 100% for group 1 in conference 2. The high percentage scores for *relevant* indicate that the participants were on topic most of the time and focused on the conference questions. The *depth* criterion ranged from a low of 78% in group 3 for conference 1 to a high of 94% for group 1 in conference 2. The high percentage score for *depth* indicates that most participants were introducing new ideas or facts. They were not simply repeating what someone else said.

This leaves the *clear* criterion. Its positive score ranged from a low of 54% for group 1 in conference 1, to a high of 75% for group 1 in conference 2. *Clear* was given a negative score if something in the analysis unit did not make sense or could not be understood. It is disturbing to see that between 25% and 46% of the analysis units presented by participants in a second-year university course had some component that could not be clearly understood.

Usefulness of the Quality of Critical Thinking Model

Maintaining consistency in the coding of the data with this model was difficult at the beginning of the coding process. Once the model was operationalized with clear examples and nonexamples, however, it became easier to use consistently. Consistently recognizing a message that would be scored as a 3 was a turning point in successfully using this model. A 3 was simply a clear expression of the participant's opinion. It was clearly stated, was relevant to the conference topic, and introduced a new idea or built on an existing idea (clear, relevant, and depth). The participant did not include any logical arguments or examples (either observations or literature sources) to support what he or she said. The level of detail or precision was low. Sometimes the 3 included *breadth* (considering an alternative explanation) rather than *depth*. A 3 became a kind of base score. A higher score generally included these three basic criteria plus one or more of the four other criteria.

In using the model it became obvious that although *clear* was a gateway, as noted by Elder and Paul (n.d), *relevant* was also a gateway. If the analysis unit was not relevant to the discussion, it received a zero for all the other criteria, because the analysis was based on the evaluation of the critical thinking about the specific conference topic. If the analysis unit was not *clear*, other criteria such as *logical* and *precise* often received negative scores, as a lack of clarity often indicated a lack of logic and precision. *Depth*, on the other hand, was less affected by a lack of clarity. *Depth* was given a positive score if a portion of the analysis unit clearly expressed an idea that was new or built on an existing idea, even if other portions of the analysis unit were not clear. *Breadth*, a criterion that looked at alternatives, was similarly generally unaffected.

The presence of a logical argument was sometimes difficult to establish at first. The Categories of Critical Thinking model was helpful here. If the analysis unit was categorized as deductive or inductive (level two of *making inferences and interpretations*), this often indicated an attempt at a logical argument. It only remained then to examine the quality of the argument.

Other criteria could be added to the model depending on the learning outcomes expected. One criterion that could have been added to this study was *concise*. Participants sometimes received a point for *precise* because they had a lot of detail in their explanations, when they might have lost one for *concise* because the explanations were much too long.

To make the model practical for use in teaching, the time required to do the coding must be reduced. Analyzing the quality of critical thinking using this model was time-consuming, even when the skills of the raters improved. Part of the difficulty was the sheer volume of material analyzed. In addition, it took some time to become proficient with the

model. One solution might be to code only a portion of the messages. This solution runs the risk of missing the messages that show the best critical thinking. Focusing on two or three criteria instead of all seven may be another solution. Beginning with *clear, relevant*, and *depth* as the base criteria and moving on to the other criteria when these are mastered or improved may be effective.

Categories of Critical Thinking Model

Using the four level-one categories of critical thinking in the categories of critical thinking model shown in Table 2, the category most frequently displayed by participants in all three groups was *making inferences and interpretations*. One of the themes in the participants' postconference questionnaires was an appreciation for expressing their own opinions. This expression of opinion may have contributed to the high number of instances of this category. The conference questions themselves promoted this type of response, as opinions were directly sought. This finding was consistent with the findings of Hara et al. (2000).

The second most frequently observed category was *clarification of the thesis, problem, or question*. The frequency ranged from 38% to 46%, a moderate score. The number of clarification statements may have been low in this study because the question was stated with a great deal of detail, resulting in few questions about the discussion topic.

Making value judgments was the least frequently observed category for all three groups, except for group 3 in conference 1. It may be that participants were reluctant to make value judgments on other participants' viewpoints or supporting evidence. One of the themes noted by all groups was an appreciation for multiple perspectives, an attitude that may have inhibited making value judgments on others' perspectives. Students are generally not experienced in judging other students' work. Moreover, they may also feel that this is not their job; rather, it is the job of the instructor. This finding is consistent with Jeong (2003), who found that students "rarely responded to arguments with evaluation of the argument's accuracy, validity, and relevancy" (p. 37). This finding is not consistent with Hara et al. (2000), who found that students did display the judgment category. The participants in their study were graduate students, which may account for their higher scores.

The next least frequently observed category was *supporting inferences and interpretations*, ranging from a low of 9% to a high of 25%, meaning that between 75% and 91% of the time participants were not supporting their statements with evidence. This is consistent with the results in the quality of critical thinking section above in which it is noted that the *support* criterion was given a negative score 75-90% of the time.

The analysis of the categories of critical thinking demonstrated by participants may provide some direction for instructors on which categories of critical thinking might require instruction. Certainly *supporting inferences and interpretations* and *making value judgments* are good candidates. It seems reasonable to expect a high percentage of the *supporting inferences and interpretations* category, because providing evidence for one's opinions is a skill expected from university level students. Most entries should present some appropriate evidence for the student's viewpoint. With percentages ranging from 9% to 25%, it is clear that few students were supporting what they said with evidence.

We do not, however, expect every entry posted by students to show *making value judgments*. There are certainly times when participants should be expressing their own opinions and not making value judgments. What percentage of the entries should be devoted to *making value judgments*? Rather than being overly concerned with how much we see of this category, should we instead be concentrating on how well it is done?

An interesting teaching technique might be to combine *making value judgments* with *supporting inferences and interpretations* by having students judge the quality of their group's entries in a computer conference based on the level of support and evidence provided by the participants in the conference.

Usefulness of the Categories of Critical Thinking Model

The original model was far too complex to be usable with its original 69 categories. As the coding proceeded, the model was simplified dramatically to 35 categories and three levels, which made coding much easier. The strength of this model is its flexibility as an assessment tool for instruction. It can be adjusted by expanding or contracting the levels to highlight specific skills that the instructor wishes to evaluate. If the focus of the instruction was clarification of the thesis, problem, or question, levels one and two of this category could be used to establish exactly which types of clarification students were and were not using. For example, students may be summarizing, questioning others, and defining terms but may not be attempting to identify and challenge assumptions. If a more general picture is required, then only level one of all four categories might be used. In this study, the *supporting inferences and interpretations* and *making value judgments* categories were demonstrated infrequently, suggesting that further work is required in these areas.

Coding the data with this model was time-consuming, mostly due to the high volume of messages. The same suggestions made for the quality of critical thinking model apply here; code only selected messages, and use only portions of the model.

Conclusion

The study highlighted quality of critical thinking skills and categories of critical thinking that were not frequently demonstrated by participants in this study. This information may be useful to instructors who are planning or designing instruction for teaching critical thinking. Quality criteria that were not frequently demonstrated included *precise*, *logical*, *support*, and *breadth*. The categories of critical thinking that were not demonstrated frequently included *supporting inferences and interpretations* and *making value judgments*.

It is hoped that the models that were adapted from other sources for this study will be useful tools in research or in teaching. Clearly the complete quality of critical thinking model and the complete categories of critical thinking model are complex and time-consuming to use, especially when analyzing large amounts of data. The strength of these models, however, is in their flexibility. The whole model is not required to conduct useful analysis. Specific criteria from the quality of critical thinking model can be isolated for instructional and evaluative use. A good place to start may be with the base criteria: *clear*, *relevant*, and *depth* (or *breadth*). The categories model can be expanded or contracted as needed up to three levels and 35 categories. Like the quality of critical thinking model, portions of the model can be selected for use depending on the analysis required.

References

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

Brookfield, S.D. (1987). *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting*. San Francisco, CA: Jossey-Bass.

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

Burt, M.T., Grady, M., & McMann, G. (1994). *Interaction analysis of an inter-university computer conference*. Paper presented at the Distance Learning Research Conference, College Station, TX.

Elder, L., & Paul, R. (n.d.). *Universal intellectual standards*. Dillon Beach, CA: Foundation for Critical Thinking. Retrieved January 11, 2000, from: http://www.criticalthinking.org/University/univlibrary/unistan.html.

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

Gabriel, M., MacDonald, C., Farres, L., & Stodel, E. (2000, May). *Exploring learner interactions in cyberspace*. Paper presented at the Canadian Association for Distance Education (CADE) Conference, Quebec City.

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.

Gibson, C. (1996). Collaborative learning in action via computer mediated conferencing. *Proceedings of the 12th annual conference on distance teaching and learning: Designing for active learning* (pp. 121-125). Madison, WI: University of Wisconsin-Madison.

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

Halpern, D.F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*, 449-455.

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

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

Henri, F. (1995). Distance learning and computer-mediated communication: Interactive, quasi-interactive or monologue? In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 145-161). New York: Springer-Verlag.

Henri, F., & Rigault, C.R. (1996). Collaborative distance learning and computer conferencing. In T.T. Liao (Ed.), *Advanced educational technology: Research issues and future potential* (pp. 45-76). New York: Springer-Verlag.

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

Kanuka, H., & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of Distance Education, 13*(1), 57-74.

King, P.M., & Kitchener, K.S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.

Lipman, M. (1991). *Thinking in education*. New York: Cambridge University Press.

McKenzie, W., & Murphy, D. (2000). "I hope this goes somewhere": Evaluation of an online discussion group. *Australian Journal of Educational Technology, 16*(3), 239-257.

McPeck, J. (1990). *Teaching critical thinking*. New York: Routledge.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mitchell, M., & Jolley, J. (2001). *Research design explained* (4th ed.). Orlando, FL: Harcourt.

Newman, D.R., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face and computer supported group learning. *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century, 3*(2), 56-77.

Paul, R. (1993). *Critical thinking: What every person needs to survive in a rapidly changing world* (3rd rev. ed.). Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R., & Elder, L. (n.d.). *Helping students assess their thinking*. Dillon, CA: Foundation for Critical Thinking. Retrieved January 11, 2000, from: http://www.criticalthinking.org/University/univlibrary/helps.nclk.

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

Rose, M. (2004). Comparing productive on-line dialogue in two group styles: Cooperative and collaborative. *American Journal of Distance Education, 18*(2), 73-88.

Cheryl McLean is the Associate Director of the Distance Education Program in the University of Manitoba. She has been with the program as an instructional designer since 1989 and served as Director from 1991 to 1996. Her current focus is the development of Web-based distance education courses. She can be reached at cmclean@ms.umanitoba.ca.