ASSESSING QUALITY OF CRITICAL THOUGHT IN ONLINE DISCUSSION

Lisa Weltzer-Ward
Walden University
Belgium

Dr Beate Baltes
Walden University
Torrance, CA
USA

Dr Laura Knight Lynn
Walden University
Des Plaines, IL
USA

Abstract

Although process, structure, and quality of online discussion have been assessed independently, a standard framework integrating these aspects for comprehensive assessment of critical thinking in online discussions is not available (Fahy, 2005; Zohar & Nemet, 2002). This paper describes a theoretically based coding framework for an integrated analysis and assessment of critical thinking in online discussion. The implementation of this critical thinking assessment framework (TAF) in future studies will ultimately help identify online educational activities and tools which best support development and application of critical thinking skills. Furthermore, it might be used to assess critical thinking of individual discussion participants or small groups. The reliability and efficacy of the framework for general discussion assessment are presented.

Introduction

Developing and utilizing critical thinking abilities is an essential aspect of education (Dewey, 1998; National, 1983). Participation in asynchronous online discussion offers the opportunity for students to develop critical thinking abilities by applying critical thinking skills, receiving feedback, and reflecting on their thinking (Baltes, 2002; Garrison et al., 2001; McKlin et al., 2001). To fully realize this learning opportunity, both assessment of critical thinking in overall discussion and assessment of individual critical thinking are required. Assessment of the overall discussion facilitates identification of best pedagogical practices such as activities that promote critical thinking, discussion prompts that solicit critical thinking, or tools which support critical thinking (Baltes, 2002; Kanuka et al., 2007). Assessment of individual critical thinking is necessary to provide feedback and facilitate reflection (Perkins & Murphy, 2006). An assessment at both the overall and individual level requires description of how the overall critical thinking process proceeds in the discussion and of how the arguments within the

discussion are structured (Fahy, 2005). Such assessment also requires analysis of quality within this process and structure (Zohar & Nemet, 2002).

This paper describes a theoretically-based coding framework for facilitating the assessment of critical thinking in online discussion by combining process, structure, and quality analysis. The integrated critical thinking assessment framework (TAF) was applied to answer the research questions, "Do students engage in critical thinking?" and "What is the quality of this critical thinking?" These questions are consistent with those which would need to be answered to assess the value of an activity, prompt, or tool for promoting critical thinking (Kanuka, 2007). This application demonstrated how the qualitative coding results from the TAF could be analyzed quantitatively to characterize the overall discussion. The application also provides reliability data for the TAF.

Development of the TAF

Dewey (1998) described critical thinking as a logical process consisting of five steps: identifying a problem, exploring the problem, suggesting a solution, judging the solution, and implementing the solution. This process proceeds through a series of highly structured logical statements which relate beliefs, data, observations, ideas, or experiences to the problem or solution and which justify judgments and conclusions. The assessment of critical thought in online discussion requires incorporating analysis of both this overall process and internal structure (Fahy, 2005). Additionally, it is necessary for online instructors to be able to assess the content quality of the contributions within that process and structure to assess learner's critical thinking skills (Fahy, 2005; Zohar & Nemet, 2002). Integration of process, structural, and quality analysis are necessary to assess critical thinking in online discussions.

An assessment model integrating process, structure, and quality could not be identified by the authors of this paper. Thus, the TAF was developed to fill this need. Research work on assessment of online discussion in each of the three areas was reviewed and then incorporated and adapted to produce the TAF. The goal was to establish a theoretically based framework which would support assessment of process, structure, and quality at the level of an individual post or statement but which could be extended to characterize overall discussion. The TAF was intended for application to a threaded, asynchronous discussion.

Dewey's five-step critical thinking model serves as a theoretical basis for research investigating the process of critical thinking in online discussion. A number of researchers have investigated what steps in the process are evident in discussion (Garrison et al., 2001; Henri, 1992; Kanuka et al., 2007; McKlin et al., 2001; Perkins & Murphy, 2006). Researchers have developed both five-step process

models like Dewey's and four-step process models in which the solutions step is integrated with either the exploration or the judgment steps. The analysis of process in online discussions showed that it is often dominated by one or two process steps and that whether or not the full critical thinking process is demonstrated depends largely upon the nature of the discussion prompt (Garrison et al., 2001; Kanuka et al., 2007; McKlin et al., 2001; Perkins & Murphy, 2006).

Other researchers have focused on the internal, logical structure of the discussion (Fahy, 2005; Jeong, 2005; Pena-Shaff & Nichols, 2004; Simosi, 2003; Zohar & Nemet, 2002). There is not a commonly accepted structural model like the steps in the process analysis, but the various analysis schemes generally include the role of statements in the discussion and how those statements are put together and related to one another. An argument framework adapted from Toulmin (1958) is commonly employed to identify the role of the statement in the discussion (Simosi, 2003; Zohar & Nemet, 2002). Claims and evidence are the central pieces of this framework. Descriptive terms such as disagreement, support, synthesis, repetition, etc. are commonly utilized to describe how statements relate to one another (Jeong, 2005; Pena-Shaff & Nichols, 2004). Analysis of the structure provides insight into how critical thought is developed and presented.

Combining analysis of the process and the structure of critical thought in online discussion provides insight into whether or not critical thought is evident in the discussion and how it proceeds (Fahy, 2005). However, this process and structural analysis does not provide insight into the quality of that critical thinking. Quality assessment also requires describing how successful the critical thinking process is and how well statements fulfill their role and relate to one another. Zohar and Nemet (2002) added a criterion-based quality assessment rubric for content knowledge to a structural analysis, and both criterion-based rubrics and indicator-based quality assessments have also been employed independently of process and structure (Newman et al., 1995; Wickersham & Dooley, 2006). However, these were criticized for having varying interpretations of criteria and indicators and inconsistencies in coding units (DeWever et al., 2006; Wickersham & Dooley, 2006).

The research findings and assessment models from these assessments of process, structure, and quality were incorporated and adapted to develop the TAF.

Description of the TAF

In order to incorporate process, structure, and quality, the TAF requires a four-step coding process for each discussion post. For the first step, the overall role of the post in the discussion process is identified. The remaining three steps are then repeated iteratively for each passage or statement in the post. In general, passages

may be parsed as sentences. However, there are cases where a sentence combines several ideas through use of conjunctions or terms like “because” or “since” linking them together. These are separated into distinct passages.

First, a post may fall into more than one of five possible identifiers (Dewey, 1998; Garrison et al., 2001; Henri, 1992):

- Initiation includes identification of a common question or problem and discussion to insure that question or problem is understood by the group.

- Exploration includes all discussion which expands upon the problem or question to support formation of a solution. Identification of appropriate assumptions to make, presentation of data relevant to the problem, and descriptions of relevant theories and facts are common examples of exploration.

- Solution includes both positing an answer or solution to the question or problem and the initial explanation of that answer or solution.

- Judgment includes all discussion where the answer or solution is debated, modified, or tested by the group.

- Resolution occurs when the participants agree upon a final solution or answer.

The second step is to identify the main role of the passage in the argument being made by the author. Then for the third step, this role and the statements relationship to other statements in the discussion is identified with a sub-code. Each passage may be given only one identifying code and one sub-code. The possible codes and sub-codes are:

- Questions pose a problem or question with the expectation of an answer within the context of the discussion (Fahy, 2005; Pena-Shaff & Nicholls, 2004). Those which elicit a single, specific answer are identified as Information. Those which elicit further discussion are identified as Discussion.

- Claims present an idea with the intention of furthering discussion. They may assert something new (Assertion), support another claim (Support), disagree with another claim (Disagreement), synthesize previous claims (Synthesis), or reply to an Information question

(Reply) (Pena-Shaff & Nicholls, 2004; Toulmin, 1958; Zohar & Nemet, 2002).

- Evidence provides specific details or information to support a claim. Evidence may be a Quotation, a paraphrased Reference, a Personal Experience of the discussion participant, an Opinion, a hypothetical or real Example, experimental Data, or a Theory which is generally accepted within the field (Fahy, 2005; Toulmin, 1958).

- Relations describe the links between claims or between a claim and evidence (Toulmin, 1958). No sub-code is applied.

- Other identifies passages that are not relevant to the discussion development. They may act in a voting nature such as "I agree" or "I don't like it" without further explanation (Scaffolding). Or, other passages may be totally unrelated showing salutations or discussing another topic (Social).

The final fourth step requires assessing the quality of all statements identified as claim or evidence using a 1 to 5 rating, ranging from 1 indicating a statement which does not show critical thinking to a 5 showing strong critical thinking and integration into an argument (DeWever et al., 2006; Newman et al., 1995; Wickersham & Dooley, 2006). Claims criteria include relevance to discussion, level of support provided for the claim, and whether relevant information is accounted for. Evidence criteria include relevance to claim, description of relation to claim, accuracy of the evidence, and source of the evidence.

Steps two through four are repeated for all statements, and the full coding process is repeated for each post.

TAF Sample Application

The TAF was applied to a discussion from a master's level online course in Philosophy of Education to answer the questions "Do students engage in critical thinking?" and "What is the quality of this critical thinking?" The instructor's prompt for the one-week long online discussion presented the 19 participants with five different general goals of education and asked them to discuss these goals and rank their importance. The goal was to assess the overall discussion, rather than individual contributions. To derive such a general description from the coded discussion, the codes for process and structure were treated as binary presence/ absence variables, and the quality rankings were treated on a 1 to 5 discrete scale. Comparisons were then made using un-paired, independent t-tests with equal

variance assumed. In all, 19 threads consisting of 43 posts were coded. In these posts, 246 distinct internal content passages were identified.

A sample coding from this application is shown in Table 1. The passages shown were a full post in the discussion. The post was coded as both solution and judgment.

Table 1: TAF Coding Example

Passage	Passage Code	Passage Subcode	Claim Rating	Evidence Reliability Rating	Evidence Relevence Rating
I would rank the goals of education in this order: #2 first,	claim	assertion	2		
because I believe, like Martha, that developing a child's self-esteem is extremely important	evidence	opinion		2	5
and also if a person is nurtured/encouraged/allowed to develop their own full, unique potential, that person is bound to make their mark as a productive, creative member of society.	evidence	opinion		2	5
And. I believe that person will be happier in the long term because they will believe that is possible to.achieve their own unique goal, whatever that may be.	evidence	opinion		2	5
The next order of importance for me would be #5,	claim	assertion	2		
and I don't accept the premise that it is utopia and will never happen.	claim	disagreement	2		
Each generation of young people can challenge the status quo and there are growing numbers of people of all generations who promote exactly those ideas of peace, harmony, equality and love.	evidence	opinion		2	4
For me, the rest of the goals would come as a matter of course if the priorities are #2 and #5.	claim	assertion	2		
In other words, if you know what your objectives are, the rest are methodologies to achieve those goals.	evidence	opinion		2	4

To address the question “Do students engage in critical thinking?” both the critical thinking process and structure were considered.

Process wise, the bulk of discussion posts, 88.6% (n = 43), presented either solutions or judgments. The full critical thought process was not evident; however,

a progression through the critical thought process is evident. Solutions dominated both early discussion (50.0%) and the initial posts in each thread (69.6%) while judgments dominated both later discussion (70.0%) and response posts in each thread (95.2%) ($p < 0.001$ for early to later and $p < 0.05$ initial to response). These indicate that participants were moving through the problem-solving process from solution to judgment as the discussion progressed. The finding that a reflective discussion prompts isolated responses to only one or two steps of the critical thinking process is consistent with other research supporting the accuracy of the TAF, and the evidence for progression provides insight beyond that previous research (Garrison et al., 2006; McKlin et al., 2001; Pena-Shaff and Nicholls, 2004).

The structure of the posts and discussion also suggested that critical thinking was evident. Almost all of the posts (97.1%) contained at least one claim, and the overall claim to evidence ratio was approximately 1.4 (119 evidence statements to 86 claim statements) indicating that claims are generally supported by at least one piece of evidence. This pattern of claim supported by evidence is indicative of critical thought (Sandoval, 2003; Zohar & Nemet, 2002).

By combining these process and structural results, we were able to identify that students did demonstrate patterns consistent with critical thinking but that they did not engage in the full critical thinking process during this reflection based discussion activity.

To address the question “What is the quality of this critical thinking?”, the quality ratings had to be considered and related to the process and structure.

The average claim ranking of 2.0 confirms that the average claim is supported by at least one piece of evidence and further shows that this evidence is generally relevant to the claim but that the claims are not supported by many pieces of evidence or by high quality evidence. This is further supported by the average ratings of 3.9 for evidence relevance but only 2.2 for evidence reliability. This showed that communication consistent with critical thinking in which the relationship between claims and evidence was generally explained but that the arguments were unconvincing or faulty. Looking at the structural sub-codes for the evidence, we were able to understand the problem with the critical thinking demonstrated. The use of opinion as evidence dominated (77.4%) the discussion. References, quotes, data, and theories which offer the most reliable support for claims were employed only 4.8% of the time. The arguments were unconvincing despite a strong claim and evidence based framework because participants in this discussion relied on opinion rather than more reliable evidence.

By adding quality rankings to our process and structural analysis, we were able to not only answer the question regarding quality but also identified not using reliable evidence as the specific problem negatively impacting the quality of student critical thinking in this discussion.

Reliability of the TAF

During the sample application of the TAF, three coders were employed to facilitate an iterative coding process during which the coding descriptions in the framework could be clarified through consensus building discussions and the overall reliability of the framework could be assessed. Table 2 summarizes the participation of each coder in the coding process and the reliability statistics for the coding. The Cohen's kappa value of 0.85 comparing coder three's independent post-based process coding to the consensus codes from coder's one and two indicated almost perfect agreement (Landis & Koch, 1977). The Cohen's kappa value 0.57 making the same comparison for the passage-based structural coding indicated moderate agreement, and the earlier comparison of the independent passage-base coding for coder's one and two resulted in a Cohen kappa value of 0.63 indicate full agreement. These values indicated that the TAF is a reliable assessment tool but that the structural framework could still be improved for better reliability.

Table 2: Coding Process and Reliability

Participating Coders			Coding Event		Percent Agreed	Cohen Kappa
A	B	C				
X	X		Independently coded 25% of discussion	process n=13	85%	0.77
				structure n=50	55%	0.37
X	X		participated in consensus building and revised discussion codes	process n=13	100%	1.00
				structure n=50	98%	0.97
X	X		Independently coded full discussion	process n=43	93%	0.77
				structure n=246	76%	0.63
X	X		participated in consensus building and revised discussion codes	process n=43	100%	1.00
				structure n=246	97%	0.95
X	X	X	discussed coding process and code descriptions			
		X	Independently coded full discussion (compared to coder A)	process n=43	91%	0.85
				structure n=246	74%	0.57

Conclusions

The TAF integrating process, structural, and quality analysis provides a powerful tool for online discussion assessment. Successful use with high reliability was demonstrated in employing the framework to answer questions about critical thinking process and quality providing information which would aid a comparison with other online discussion activities or with a similar discussion using support tools. The proportions of process steps in the discussion, claim to evidence ratio, claim rating, evidence reliability rating, and evidence relevance rating provide clear metrics to use for such a comparison. This simple example application also demonstrated the further power of this framework when applied in the classroom. The problem of using opinions rather than reliable evidence was specifically identified. Knowing this problem, an intervention could be initiated to help students with this specific aspect of critical thinking. Although not demonstrated here, such analysis to identify needs could be made at the individual student level as well. Information regarding participation in various aspects of the process, types of claims or evidence used, quality of those claims and evidence, and overall argument structure could be used to determine the strongest and weakest aspects of a student's critical thinking. The TAF could also be combined with content analysis to support research like that of Wade, Fauske, and Thompson (2008) in which the theories and content applied by participants is considered in conjunction with their critical thought development.

This work represents initial development of the TAF. Further work will focus on application of the framework, development of the framework for improved clarity and reliability, and integration of the framework with other aspects of discussion assessment such as participation and subject specific content analysis.

References

Baltes, B. (2002). Virtual classroom discussions versus traditional classroom discussions. In H. X. Kinshuk (Ed.). *Learning communities on the Internet: Pedagogy in implementation*. Los Alamitos, CA: IEEE Computer Society.

DeWever, B., Schellens, T., Valcke, M., & VanKeer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers and Education, 46*(1), 6–28.

Dewey, J., Hickman, L. A., & Alexander, T. M. (Eds.). (1998). *The essential Dewey: Volume 2 ethics, logic, psychology*. Bloomington, IN: Indiana University Press.

Fahy, P. J. (2005). Two methods for assessing critical thinking in computer-mediated communications (CMC) transcripts. *International Journal of Instructional Technology and Distance Education, 2*(3), Article 02.

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

Henri, F. (1992). Computer conferencing and content analysis. In A. Kaye (Ed.), *Collaborative learning through computer conferencing, the Najaden papers Vol. 90* (pp. 117–136). Berlin, Germany: Springer-Verlag.

Jeong, A. C. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367.

Kanuka, H., Rourke, L., & Lallamme, E. (2007). The influence of instructional methods on the quality of online discussion. *British Journal of Educational Technology, 38*(2), 260–271.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 321–325.

McKlin, T., Harmon, S. W., Evans, W., & Jones, M. G. (2001). Cognitive presence in web-based learning: A content analysis of student's online discussions. *Annual Proceedings of Selected Research and Development Practice Papers Presented at the National Convention of the Association for Educational Communications and Technology, 24*(1–2), 272–277.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Retrieved March 12, 2008, from http://www.ed.gov/pubs/NatAtRisk/index.html

Newman, D. R., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face and computer supported learning. *Interpersonal Computing and Technology, 3*, 56–77.

Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussion. *Computers and Education, 42*(3), 243–265.

Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Education Technology & Society, 9*(1), 298–307.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12*(1), 5–51.

Simosi, M. (2003). Using Toulmin's framework for the analysis of everyday argumentation: Some methodological considerations. *Argumentation, 17*, 185–202.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Wade, S. E., Fauske, J. R., & Thompson, A. (2008). Prospective teachers' problem solving in online peer-led dialogues. *American Educational Research Journal, 45*(2), 398–442.

Wickersham, L. E., & Dooley, K. E. (2006). A content analysis of critical thinking skills as an indicator of quality of online discussion in virtual learning communities. *Quarterly Review of Distance Education, 7*(2), 185–193.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62.