Further Classification and Methodological Considerations of Evaluations for Online Discussion in Instructional Settings

Alexandru Spatariu  
Houston Baptist University, United States  
aspatariu@hbu.edu

Denise L. Winsor  
University of Memphis, United States  
dwinsor@memphis.edu

Cynthia Simpson  
Houston Baptist University, United States  
csimpson@hbu.edu

Eric Hosman  
University of Memphis, United States  
ehosman@memphis.edu

ABSTRACT
With the rapid advancements of technology, online communication in both K-12 and post-secondary instruction has been widely implemented. Instructors as well as researchers have used various frameworks to evaluate different aspects of online discussions’ quality. The online discussions take place synchronously or asynchronously in chat rooms, boards, and blogs, often using mobile applications and usually aimed at understanding course content and concepts. The current review follows up on Spatariu, Hartley, and Bendixen’s (2004) classification that placed these frameworks in four categories based on what they were aimed at measuring (disagreement, argumentation, interaction, and content). The current review serves two main purposes. First, newer frameworks are categorized and described while addressing methodological considerations. Second, conclusions and recommendations for future research and instructional applications of online discussion evaluation are made.

INTRODUCTION
A report by two research groups that are tracking distance education yearly in the United States (Allen & Seaman, 2013) shows that there were 6.7 million students enrolled in higher education online courses in 2011. Straunshiem (2014) reported that about 2.6 million students were enrolled in fully online programs while the rest were taking some online courses. Graduate students are typically the ones who opt for completely online programs rather than undergraduate students (22% versus 11%). While higher education has slowed its expansion in the last few years, K-12 education has been rapidly increasing. North American Council for Online Learning (2012) reports 26 states have state virtual schools, 31 states and Washington, DC have state-wide full-time virtual schools with an estimated total enrollment of 1.8 million students in 2009-2010. The delivery mode in K-12 education has also been summarized by NCES (2012) with 53% of public high schools reporting 1.3 million students enrolled in distance educations courses in 2010. Keeping these educational trends as well as the rapid progress of technology in mind, one can surmise all aspects of distance education have to be continuously researched and improved, including online discussions and communication.

Online discussions, also known as online discourse or computer mediated communication, can be synchronous (e.g., chat rooms) or asynchronous (e.g., discussion boards) and are common practice in many types of distance education courses. Online discourse is used for purposes such as understanding subject matter, enhancing communication, developing cooperative projects, and boosting critical thinking skills (Bonk & Dennen, 2007; Garrison, Anderson, & Archer, 2000, 2001; Kay 2006; Meyer 2003; Palloff & Pratt, 2001; Rourke & Anderson, 2002; Spatariu, Quinn, & Hartley 2007; Spatariu, Hartley, Schraw, Bendixen, & Quinn, 2007; Tu & McIsaac, 2002).

In order to evaluate the quality of online discourse when using either course-based online discussion tools (e.g., discussion boards, chat) or similar tools ancillary to the course (e.g., wikis, skype, mobile device applications) different frameworks have been employed. A framework is a grading rubric that allows the reader to score the discussion (e.g., interactivity patterns, strength of an argument). Spatariu, Hartley, and Bendixen (2004) classified and described a number of such frameworks, placing them in four categories based on the constructs...
that were purportedly measured by the instructors. The categories were levels of disagreement, argument structure analysis, quality of interactions, and content analysis. These frameworks provide a foundation for researchers and practitioners interested in a systematic and purposeful way of evaluating the quality of course discussion as it relates to course objectives or goals.

The current review follows-up on the frameworks presented in Spatariu et al. (2004) and explores new frameworks. It also discusses methodological considerations and provides suggestions for future use. First, the conclusions of Spatariu et al. (2004) are reviewed to illustrate specific evaluation models. Second, new frameworks are reviewed that pertain to evaluation of argumentation, interaction, content, and qualitative analysis. Extensive literature searches were conducted to locate evaluations frameworks employed in research studies, especially those published in the past 5-6 years. Particular information, related to the type of study, theoretical framework, and reported reliability and validity undertakings, is included in three different tables. Many studies, even though recently published, were not included in this review as the overall focus was on number of instructor or student posts, replies, time, length, and other descriptive features of the generated discussions. While of possible value to research, this type of information was not considered to be particularly relevant to the quality of the actual discourse. The focus of this review was on studies that involved substantial analysis of the writing involved within discussions. Lastly, conclusions and recommendations for future research and practice for discourse in both post-secondary and K-12 instruction are presented.

EXISTING FRAMEWORKS

Levels of disagreement and argument structure analysis are approaches that have been used by different researchers (Golanics & Nussbaum, 2008; Spatariu et al. 2007) to evaluate the quality of arguments produced in online discourse. Although their coding schemes vary based on research needs, they all targeted agreements, disagreements, and evidence supplied in support of claims. At a basic level, argument and counter-arguments can be counted and recorded. At an advanced level, the type of claim and evidence would make an argument weak or strong, and would allow the reader to score it beyond simple categorization as agreement and/or opposition.

Interaction based coding has been used by other researchers such as Schaeffer, McGrady, Bhargava, and Engel (2002), Järvellä and Hakkinen (2002), and Nurmela, Lehtinen, and Palonen (1999). The main purpose of these methodologies is to identify particular message roles in the larger discussion. Message board posts are usually scored based on the relationships they establish with other posts, especially as related to perspective-taking, change of topic, and type of social interaction.

Spatariu et al.’s (2004) research included the last category, content analysis. Several studies (e.g., Hara, Bonk, & Angeli 2000; Henri, 1992; Peterson-Lewinson 2002) have developed frameworks that examine such learning aspects as cognitive and metacognitive skills and depth of processing, as well as social interaction and participation patterns.

NEW FRAMEWORK: ARGUMENTATION ANALYSIS

Researchers continue to further develop and use argument structure analysis frameworks. Clark and Sampson (2008) developed and employed an analytic framework for assessing argumentation in online science courses that examined levels of opposition, discourse patterns, use of evidence, and conceptual soundness. They have also reported on validity and reliability of the instrument. Salminen, Marttunen, and Laurinen (2010) have embedded argumentative discourse in chat discussions. This approach was quite different from other asynchronous argument analysis frameworks as students had the opportunity to construct argument diagrams with or without computer assistance. The diagrams produced were analyzed for different argument structures and inclusion of prior knowledge.

Other researchers such as Clark, Samson, Weinberger, and Erkens (2007) examined methodological aspects of existing frameworks for argument structure analysis. Their review looked at argument structure and conceptual quality, which exist in most frameworks presented. Their work explores aspects of previous argumentation analysis frameworks employed by Clark and Sampson (2008) in their study, which is included in the table below. Additionally, researchers have employed various evaluation schemes that included evaluation of arguments along with other types of post characteristics such as elicitation and integration (Tawfik, Sánchez, & Saova, 2014).
### Table 1: Argumentation Analysis Frameworks

<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Framework</th>
<th>Theoretical Framework</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clark &amp; Sampson (2008)</td>
<td>Argumentation in asynchronous discussions</td>
<td>-Dialogic arguments to reach agreements on ill-defined problems</td>
<td>Inter-rater reliability 94% (Cohen’s k = 0.91)</td>
<td>Framework scores the individual comments in terms of discourse moves, grounds quality, &amp; conceptual quality -The framework is based on previous frameworks; each modification is discussed and justified</td>
</tr>
<tr>
<td>Salminen, et al. (2010)</td>
<td>Argumentation in synchronous chat discussions</td>
<td>Three theories were discussed as they pertain to the use of visual argument diagram construction: the theory of computational efficiency, the cognitive theory of multimedia learning, and the cognitive load theory</td>
<td>Not reported</td>
<td>Framework is based on participants constructing visual argument diagrams -Participant-generated diagrams were compared and classified based on categories supported by previous research</td>
</tr>
</tbody>
</table>

**NEW FRAMEWORK: INTERACTION ANALYSIS**

Recently research has adopted and further developed a social interaction analysis framework. However, the social interaction framework is not mutually exclusive with the community of inquiry framework which suggests that there is overlap in what they propose to evaluate in the discourse.

Hull and Saxon (2009) evaluated the social interaction of education courses during asynchronous discussions. The evaluation instrument has been previously used and focused on the presence of thought process patterns in discussions, in addition to evaluation and explanation of social, cognitive, and metacognitive processes detected. Hull and Saxon (2009) detected higher mental processes and more sophisticated interaction patterns than previous frameworks, which may mean the evaluation framework they employed is more elaborated. Heo, Lim, and Kim (2010) employed both social network analysis and content analysis to evaluate levels of interaction and knowledge construction in project-based learning environments. The authors neglected to investigate methodological issues of the instrument most likely because it was based on a previously developed and tested framework. However, they concluded the tool needs further development to address emerging coding (qualitative analysis codes not previously classified, which surface while analyzing data). Likewise, Lang (2010) examined interaction in project-based learning environments at the high school level using asynchronous discussions. This evaluation of discourse focused on information exchange, knowledge construction and negotiation. The findings of Heo, et al. (2010) and Lang (2010) are based on the framework developed by Gunawardena, Lowe, and Anderson (1997) for measuring social interaction patterns. Although there is valuable information about turn taking and conversation patterns that these frameworks can provide, the overall trend is to develop evaluation tools that get more extensively into what is being discussed, what type of reasoning is involved, and how deeper thinking is manifested. The need for more complete understanding of participants’ thinking and interactivity has led some researchers such as Heo et al. (2010) to employ two different frameworks, in their case both social interaction and content analysis.

### Table 2: Interaction Analysis Frameworks

<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Framework</th>
<th>Theoretical Framework</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
</table>
| Hull & Saxon (2009) | Social interaction in asynchronous discussions | -Social construction of knowledge -Social collaboration | Inter-rater reliability (k=0.77) | -Framework is based on previously developed frameworks for social interaction and knowledge construction -Coding included the following categories: direct instruction, sharing new information, situated definition, inter-subjectivity, negotiation/co-
### NEW FRAMEWORK: CONTENT ANALYSIS

An important and fairly large body of research, that includes but is not limited to coding and analysis of discussion transcripts, has been initiated in the work of Garrison, Anderson and Archer (2000) who coined the term *community of inquiry*. Their work stems from Henri’s (1992) content analysis work, but they created a comprehensive instrument for the description and analysis of the online-environment educational experience consisting of three main elements: social presence, cognitive presence, and teaching presence (Garrison et al. 2000). Numerous subsequent studies (Cleveland-Innes, Koole, & Kappelman, 2006; Garrison, et al., 2001; Garrison, Cleveland-Innes, & Fung, 2004; Gorsky, Caspi, Antonovsky, Blau, & Mansur 2010) have employed this model to evaluate the three components and their particular descriptors: social presence (i.e. expression, group cohesion), cognitive presence (i.e. resolution, integration) and teaching presence (i.e. type of instructor involvement, shifts in presence). This framework has been employed in a variety of courses for content transcript analysis to include problem-based learning in agriculture (Kenny, Bullen, & Loftus 2006), natural sciences and humanities (Gorsky, et al. 2010), teacher education (Koh, Herring, & Hew 2010); and English language (Ho & Swan 2007).

Other researchers have adopted the Garrison et al. (2004) community of inquiry framework explain the community of inquiry framework. Tirado, Hernando, and Aguaded (2012) and others have employed framework combinations; for instance, Tirado et al. (2012) used a combination of content analysis as initiated by Henri (1992) and social network analysis as used by Wang and Li (2007) and Reffay and Chanier (2002). These combination frameworks tend to be focused on social presence and cognitive presence factors.

Shea, et al., (2011) used both the community of inquiry framework and learning outcomes taxonomy to evaluate online asynchronous discourse. Aykol and Garrison (2011b) employed transcript analysis to assess cognitive presence in both online and blended communities of learning. Results revealed students achieved high levels of cognitive presence and learning outcomes. Aykol and Garrison (2011a) further developed content analysis into a metacognition evaluation instrument. The community of inquiry theoretical framework served as a conceptual base for metacognitive constructs, operationalization, and evaluation. The use of content analysis, just like many other frameworks, has been employed in chat discourse analysis (Hou & Wu 2011). Another social analysis framework, discourse analysis, was employed by Dennen and Wieland (2007) and by Herring (2004). Discourse analysis consisted of scoring social engagements, acknowledgments, peer questioning, and perspective taking. There are many overlaps of this framework with both argumentation and interaction frameworks, which have already been discussed. Joreczak and Bart (2009) also employed a framework that evaluates both cognitive structures, through content analysis, and argumentation patterns in asynchronous discussions.

Kay (2006) presented a comprehensive framework for analyzing the quality of online discussions. This framework stems from content analysis (Hara et al. 2000) and the social aspects of learning (Vygotsky 1978). Some of the variables measured included aspects of social learning, cognitive involvement, discussion structure, instructor role, discourse challenges, learner attitudes, and learning performance. Putman, Ford, and Tancock (2012) developed their own framework for collaboration and cognitive engagement based on students’ discourse data.

Another approach for cognitive presence evaluation is based on Bloom’s taxonomy (Valcke, De Wever, Zhu, & Deed, 2009). A unique aspect of this study is that the authors did not use a learning management system designed for online courses; instead they utilized social media (i.e., Facebook) as the interaction space for a
project-based learning activity. Their instruments detected both low level cognition (i.e., understanding and comprehension) and metacognitive processes. Higher order thinking skills were examined by Xie and Bradshaw (2008) as well in an experimental study on the effects of questioning prompts on solving ill-structured problems. The authors developed their own coding scheme that was essentially a rubric for detecting identification and possible solutions of the various problems presented for discussion. Problem identification and solution each contained four criteria related to number of problems, justification of problem, number of solutions, justification of solution, quality of solution, etc. Two raters scored the students’ posts to ensure reliability. A similar rubric was designed to evaluate problem-solving abilities in a study by Du, Yu, and Olinzock (2011). They looked at the effects of instructor prompts on different types of discourse from chat rooms to discussion boards, and evaluated the assignments using rubrics that yielded significant differences on problem construction, needs assessment, and argument construction.

<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Framework</th>
<th>Theoretical Framework</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorsky, et al. (2010)</td>
<td>Teaching, cognitive, and social presence in asynchronous discussions; content analysis</td>
<td>Community of inquiry</td>
<td>Inter-rater reliability at 92% (Cohen’s k=0.89)</td>
<td>Validity is discussed based on validity reported for previously developed framework upon which the current one is based</td>
</tr>
<tr>
<td>Koh, et al. (2010)</td>
<td>Teaching, cognitive, and social presence in project-based learning asynchronous discussions; content analysis</td>
<td>Community of inquiry</td>
<td>Inter-rater reliability (k=0.75)</td>
<td>Framework based on previously developed codes related to knowledge construction, teaching, social interaction, and logistics</td>
</tr>
<tr>
<td>Tirado, et al. (2012)</td>
<td>Social interaction and cognitive presence in asynchronous discussions</td>
<td>Community of inquiry</td>
<td>Triangulation of data used for reliability</td>
<td>Validity is discussed based on existing content and social network analysis frameworks</td>
</tr>
<tr>
<td>Shea, et al. (2011)</td>
<td>- Teaching, cognitive, and social presence -Learning outcomes taxonomy</td>
<td>Community of inquiry</td>
<td>Inter-rater reliability using Holstí’s Coefficient of Reliability</td>
<td>Validity is discussed based on existing frameworks</td>
</tr>
<tr>
<td>Aykol &amp; Garrison (2011b)</td>
<td>-Cognitive presence -Learning outcomes -Content analysis</td>
<td>Community of inquiry</td>
<td>Inter-rater reliability at 75%</td>
<td>Validity is discussed based on collection and analysis of different types of data</td>
</tr>
<tr>
<td>Hou &amp; Wu (2011)</td>
<td>-Content analysis -Lag sequential analysis in synchronous discussions</td>
<td>Social learning</td>
<td>Inter-rater reliability (k=0.67)</td>
<td>Validity is discussed based on existing frameworks</td>
</tr>
<tr>
<td>Aykol &amp; Garrison (2011a)</td>
<td>Metacognition in asynchronous discussions</td>
<td>Community of inquiry</td>
<td>Not reported but discussed</td>
<td>Discussed based on existing metacognition constructs and instruments</td>
</tr>
<tr>
<td>Valcke, et al. (2009)</td>
<td>- Cognitive processing categories in Bloom's taxonomy - Cognitive, affective, and Social interaction</td>
<td></td>
<td>Inter-rater reliability reported for both instruments (and sections of the instruments) ranging from K=0.87 to 0.95</td>
<td>Not explicitly discussed but instruments are based on existing constructs that are discussed</td>
</tr>
</tbody>
</table>
metacognitive learning

| Xie & Bradshaw (2008) | Solving ill-structured problems, critical thinking | -Collaborative inquiry | Inter-rater reliability represented by Pearson correlation reported on problem representation 1 ($r = .856$, $p < .001$), representation 2 ($r = .745$, $p < .001$), representation 3 ($r = .738$, $p < .001$), and representation 4 ($r = .821$, $p < .001$). And on problem solution 1 ($r = .698$, $p < .001$), solution 2 ($r = .756$, $p < .001$), solution 3 ($r = .781$, $p < .001$), and solution 4 ($r = .811$, $p < .001$). | Scoring rubric is based on an existing instrument; additionally two experts in the field of educational psychology reviewed the rubrics prior to implementation in scoring. |

NEW FRAMEWORK: QUALITATIVE ANALYSIS

Some researchers use qualitative approaches to evaluate online discourse. An advantage of a qualitative approach is the possibility of exploring new aspects of discourse that may not be captured in a previously constructed framework. For example, Rourke and Kanuka (2007) incorporated a unique approach to online discussion evaluation in which they conducted post-qualitative analysis and interviewed students about their interactions and writing experiences. Other researchers examined the level of critical thinking and involvement of students in asynchronous discussions (Lim, Cheung, & Hew 2011; Vonderwell, Liang, & Alderman 2007). This approach yielded information on student exchange of information that may not have been adequately captured by an existing framework that quantified the information of messages.

Arend (2009) used a mixed methods approach to explore critical thinking patterns in online asynchronous discussions. The emphasis of this particular study was on qualitative analysis that revealed many subtle aspects of advanced critical thinking when instructor involvement is more purposeful and less prevalent. Baran and Correia (2009) employed basic quantitative approaches (number of posts, type of posts) and qualitative approaches (discourse evaluation) in mini case studies to analyze students’ discussions in education classes. They also used triangulation of discourse data, course materials and instructor guidelines to strengthen the study’s trustworthiness. Findings of the study suggest student-led discussions can be very instrumental in boosting motivation to participate in discussions, generation of new ideas, and the creation of an environment conducive to overall learning.

In summation, qualitative approaches allow for exploration of new discourse aspects that may not be otherwise captured when employing an evaluation tool already in use. However, in some cases, constructs purportedly being explored in these qualitative studies have many similarities with existing frameworks previously described and that would have to be investigated by the researcher before using in online discussion analysis.

RECOMMENDATIONS

The current paper updates Spatariu et al.’s (2004) review to provide an overview and evaluation of the newer frameworks for evaluating different aspects of quality in online discussions. Studies were placed in four categories of analysis: argumentation, interaction, content, and qualitative. The classification is primarily for the ease of understanding the concepts targeted for measurement, although there are areas of overlap. An important aspect of choosing one approach over another for research or practical reasons involves considering both discussion implementation (i.e. accomplishing course goals) and the evaluation of the discourse (i.e., grading, instrument validation).

METHODOLOGICAL CONSIDERATIONS

Below we discuss a few methodological aspects that can help in advancing research in this field. It is important to note that some of the instruments presented need additional testing for validity and reliability. There is a substantial amount of research moving in this direction for some of the frameworks presented, while others are
isolated studies that cannot claim sound generalizability based on quality measurement. For example, community of inquiry has received a lot of attention in the literature and some articles examined validity and reliability evidence (Garrison, 2007; Garrison et al. 2004; Garrison et al. 2006). Further, DeWever, Schellens, Vackle, and Keer (2006) examined 15 content analysis frameworks for evaluating online discourse. They paid particular attention to the theoretical base, validity and reliability reporting, and the choice of the unit of analysis. As the three tables illustrate, some of the newer frameworks provide the reader with information on validity and reliability (Akyol & Garrison, 2011b; Heo et al., 2010; Hou & Wu, 2011; Hull & Saxon, 2009; Shea et al., 2011) while others suggest more studies need to be conducted (Akyol & Garrison, 2011a; Salmimen et al., 2010). It appears as though newer analytical frameworks are grounded in particular learning theories.

Penny and Murphy (2009) took a different, more practical approach; they collected, compared, and analyzed 50 rubrics being utilized for college level asynchronous discussion evaluation. They studied the commonalities among these rubrics and placed them in the following categories: cognitive, mechanical, procedural and interactive. This type of research and analysis can be useful for practical applications; however, we encourage more in-depth exploration of each instrument’s methodological issues. For example, Rourke and Kanuka (2009) conducted a comprehensive literature search of over 250 articles that involve community of inquiry and reported that only five of them included a concrete measure of student learning. This means that no validity evidence was advanced indicating the method accurately and consistently measured student learning outcomes.

It is important that future research considers other salient aspects when examining online discussion quality, for example, accuracy, time requirements, and trainer scoring issues (Meyer, 2003). We suggest further work should be done in automated computerized assessment systems based on these frameworks. Some researchers have already developed tools along these lines such as the discussion analysis tool (Jeong 2003; Jeong, Clark, Sampson, & Meneke 2011). However, more research is needed to improve the operation, functionality and performance of computerized assessment systems, as they can be difficult to learn how to use.

Lastly, more research needs to be conducted to determine how the current constructs measured by these frameworks correspond to other learner characteristics such as motivation (Zhang, Koheler, & Spatariu, 2009), metacognition (Hou & Wu, 2011), and epistemology (Nussbaum, Sinatra, & Poliquin, 2008). One way to show evidence of construct validity is through looking at other constructs (convergence) to see how they are related to discourse frameworks. Zhang et al. used a unique approach to identify some of the more outlying learner characteristics by developing and validating an instrument for motivation for critical reasoning in online discourse. This type of instrumentation can provide data on how motivation for reasoning is related to argumentative aspects of online discussions or higher levels of critical thinking as exhibited in online discourse. Hartnett (2012) conducted research that reveals the importance and complexity of relationships between motivation, participation, and achievement of pre-service teachers in online asynchronous discussions. Both holistic learner approaches as well as particular constructs related to learning approaches have to be further developed and explored to further the field’s understanding of ways to analyze online discussions quality.

REFERENCES


Copyright © The Turkish Online Journal of Educational Technology


