Building the SCAFFOLD for Evaluating Threaded Discussion Forum Activity: Describing and Categorizing Contributions

Laurie P. Dringus and Timothy J. Ellis

Abstract - As asynchronous learning networks mature, a demand has been placed on incorporating “presence” and interaction in the online learning experience. The threaded discussion forum is a promising asynchronous tool for promoting the desired student and faculty interaction essential to fostering a collaborative community of learners.

Although the threaded discussion forum is widely used to support a variety of course requirements, the instructor is faced with the difficulty of interpreting and evaluating the learning and quality of participation reflected in the student contributions. A considerable discussion in the literature is ongoing on the theoretical foundation for discourse analysis of discussion forum activity. There have also been a number of attempts to develop rubrics for evaluating that activity. Unfortunately, many of the rubrics have little or no grounding in discourse analysis theory, and those that are conceptually sound are either too cumbersome for a professor to use or have not been tested for reliability and validity.

The presentation will describe SCAFFOLD – Scale for Forums/Online Discussion Assessment. The authors will describe the SCAFFOLD instrument as a tool for categorizing and describing contributions. The process for establishing reliability and validity will be detailed.

Index Terms - asynchronous discussion forums, distance education, online learning environments, online pedagogy.

INTRODUCTION

Perhaps one of the most challenging obstacles to overcome in an asynchronous environment is how to promote the sense of “presence” and “community” among participants, both students and faculty. Advances in bandwidth capacity and video compression effectiveness are beginning to expand promise of providing potentially more powerful solutions to the difficulty of promoting a sense of presence in an asynchronous environment. One asynchronous tool in particular – the threaded discussion forum – remains one of the most widely accepted mechanisms for affording high level interaction in various dimensions, including student-to-student, student-to-teacher, and teacher-to-student interaction. High level communication is necessary for fostering a community of learners.

Although the threaded discussion forum is widely used to support a variety of course requirements, the instructor is faced with the difficulty of interpreting and evaluating the learning and quality of participation reflected in the student contributions. When postings to a single topic can run into the hundreds, span several weeks, and represent many different types of contribution to the discourse, the view of an individual student’s experience in the learning activity is not clear. The student’s success is represented by the content of the individual contributions posted and the context in which the postings were made. The sheer volume of postings makes it difficult, if not impossible, to analyze content and context for each student. The instructor is faced with the dilemma of having access to a great amount of data that could potentially be of use in evaluating a student’s performance but has limited capacity for processing that data into meaningful information.

There has been considerable discussion in the literature on the theoretical foundation for discourse analysis of discussion forum activity. There have also been a number of attempts to develop rubrics for evaluating that activity. Unfortunately, many of the rubrics have little or no grounding in discourse analysis theory, and those that are conceptually sound are either too cumbersome for a professor to use or have not been tested for reliability and validity.

The goal of this study was to build upon the theoretical foundation established in the literature to develop a meaningful, usable schema for categorizing and describing discussion forum contributions. The result of this effort is SCAFFOLD – Scale for Forums/Online Discussion Assessment – a tool for categorizing and describing contributions in threaded discussion forums.

The balance of this paper is organized as follows. A brief discussion of the theory and research underlying discourse analysis as it applies to threaded discussion forum assessment is followed by a detailed discussion of the methodology utilized in developing SCAFFOLD. The article concludes with an analysis of implications for future research directions. It is important to note that SCAFFOLD is in its preliminary stage of development. The results presented involve an initial test of...
the methodology that the authors desire to refine in future steps towards validation.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>PARTICIPATION INDICATORS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participation Construct</strong></td>
<td><strong>Example Indicators</strong></td>
</tr>
</tbody>
</table>
| Level of interaction in forum; Learner-learner interaction activity [7] | • Level of discussion – high, progressive, low level [10]  
• Message type: messages, contributions, responses, and general postings  
• Amount of interaction [11] |
| Degree of “presence” in forum; Cognitive and Social presence [2] | • Early, middle, late, or last minute contributions  
• Critical thinking skills – creativity, problem solving, intuition, insight [2]  
• Practical inquiry – triggering event, exploration, integration, resolution [3]  
• Encouraging collaboration among peers [2] |
| Timing and pace: respond to others in a timely fashion | • Instructor’s definition of the immediate vs. latent continuum  
• Interval “wait” time [12]  
• For responses, time between initial posting and response |
| Staying on topic; Learner-content [7] | • Meaningful and relevant keywords, phrases used to stay on topic |
| Transitions; Turn taking activity; Shifts in topical focus | • Transitions/continuous discussion  
• Interval “wait” time [12]  
• Initiating versus responding to topics [13]  
• Within/outside the boundaries of the thread domain; where the changes occurred; subtle and dramatic shifts |
| Extent of instructor interaction; facilitation; Instructor-learner activity [7]; Teaching presence [2] | • Facilitation rating– highly facilitative, informative, useful, non-facilitative  
• Instructor lag time; feedback [14]  
• Defining and initiating discussion topics; focusing discussion [2]  
• Instructor-learner knowledge building processes [5] |
| Mandatory/non-mandatory participation | • Lag time between postings  
• Increased or decreased interaction and mandatory or non-mandatory directives about topics and group size [14] |
| Lurking [8]; Vicarious interaction [9] | • Time spent on reading, writing, composing messages  
• Extent of reviewing vs. participating [8]  
• Extensive/exclusive use of summarizing postings  
• Absence/lag time between postings |
| Shared resources | • Citations from scholarly literature  
• Usable resources such as links to informative websites |
| Accuracy of message content | • Content is accurate, valid, and relevant; writing quality |
| Accuracy response placement | • Responses are placed in the proper sequence |
| Contribution to “group process” (Gestalt view); group size | • Quality of contributions/the number of contributions made; conflict/negotiation: agreements/disagreements made in responses [4][15]  
• Position statements versus new points  
• Group think versus individual effort  
• Group size and proportion of interaction [14] |

**Evaluating Threaded Discussion Forums**

Researchers have developed models for analyzing the process of learning in asynchronous computer conferencing. Henri’s [1] *Analytical Model* and Garrison, Anderson, and Archer’s [2]-[3] *Practical Inquiry Model of Cognitive Presence*, are extensive content analysis models partially based on discourse theory, cognitive theory, and interaction theory. Garrison et al. [3] and Jeong [4] provided an invaluable assessment tool for analyzing learning and interaction constructs reflected in text-based computer conferencing transcripts. One key recommendation that Garrison, et al. [3] made was that there is a need to develop tools that effectively manage large numbers of messages in longer running online courses. Schrire’s [5] follow up study to Garrison, et al.’s [3] work also indicated that the coding process relies heavily on manual interpretation by the human rater, suggesting that more formidable content analysis tools should be embedded within the threaded discussion forums assessment toolset to automate coding processes. Roblyer and Wiencke [6] provided a rubric for assessing interactive qualities in distance courses. Their rubric serves as “a tool to allow for more meaningful examination of the role of interaction in enhancing achievement and student satisfaction in distance learning courses” (p.77, abstract).
The research related to assessing student interactivity in computer conferencing and asynchronous discussion forums continues to expand to provide new techniques that could be used as standard assessment practice. In many cases these works [3], [1], [4], [6] and others, point to the need for assessment tools to reduce the cumbersome manual assessment process that burdens the instructor.

**TABLE II**

**SCAFFOLD Version 1**

**SCALE FOR FORUMS/ONLINE DISCUSSION ASSESSMENT:**
Evaluating Content and Role of Contribution in Online Course Discussions

How well does each of the following statements describe the contribution? Please keep in mind:
1. Few contributions would likely contain evidence of all indicators
2. Your goal is to categorize or **describe the contribution**, not to evaluate its value
3. You will be using a 3 point scale in categorizing each contribution on which a rating of 1 equals “Not Evident”, 2 equals Somewhat Evident, and 3 equals “Strongly Evident”.


Table 1 provides a sample list of participation constructs and associated indicators that are commonly used by instructors to assess forums. Many of these are often undefined or are used in different contexts, making it difficult to appropriately categorize and match patterns of constructs and indicators. In Table 1, literature citations are noted when key constructs, phrases, or indicators have stood out in current research. The table is a start for illustrating the need for a comprehensive list of participation constructs and indicators to be addressed in future work.

The most common and often cited (Table 1, Column 1) participation constructs include levels of interaction in the forum (most commonly referred to as “learner-learner, learner-content, instructor-learner interactivity” by Moore [7], degree of presence (most recently specified by Garrison, et al. [2][3], in context of critical thinking and in terms of cognitive, social, and teacher presence), and lurking [8] or vicarious interaction [9]. Other participation constructs (i.e., those not cited with literature in Table 1) have had a long history in that they are often commonly accepted among those discussing online participation in various contexts. For example, participation constructs such as timing and pace, transitions, facilitation, shared resources, accuracy of postings, and contributions to group process are often used as common terms in the literature.
The range of participation indicators (Table 1, Column 2) – those key phrases or terms that are used to represent “data” to assess the various constructs – are more likely identified in piecemeal in the literature and used to discuss specific contexts of participation. For example, some participation indicators have been noted in the literature including graphing levels of discussions in categories, such as high level discussions, progressive discussions, and low-level discussions [10]; amount of interaction [11]; critical thinking skills and practical inquiry transactions [2][3]; interval “wait” time [12]; instructor lag time and feedback [14]; and indicators related to group process and group size [14], [15], [4].

**Methodology**

The literature summarized in Table 1 resulted in a list of 19 participation indicators (Table 2). It is intuitively evident that a schema consisting of 19 separate points would not be very useable in analyzing a large number of forum contributions; further, it is doubtful that the very large volume of data produced by such an analysis would be meaningful. A much more compact but still theoretically sound schema was indicated. Toward the end of constructing that more manageable instrument, an initial version of SCAFFOLD comprised of all 19 points was used in a pilot study in order to obtain a preliminary set of data upon which an initial factorial analysis could be conducted. The authors desired to see the patterns of groupings of indicators that would result from the initial use of SCAFFOLD with 19 indicators. The following sections detail the assumptions, limitations, and delimitations underlying the study, the approach followed in the pilot study, and the statistical tests used to analyze the data resulting from that study.

**Assumptions, Limitations, and Delimitations**

This study must be reviewed in the context of two major assumptions: the raters who participated in the pilot study were a representative sample of college-level instructors who might potentially teach an asynchronous course, and the threaded discussion forum contributions assessed using the SCAFFOLD instrument were representative of discussion forum contributions that would be encountered in a college-level course. The validity of the results of this study may be influenced by several limitations. One limitation was that the raters (n=27) who participated in the study did so as members of a doctoral-level course in which they studied the research area of Online Learning Environments under the direction of the first author. Many of the raters were experienced professors, with approximately 25%, as identified as having taught or currently teaching online courses at their institutions. It was not possible to clearly identify or predict the influence that their “student status” had on the manner in which they assessed the contributions. Another limitation was that rater fatigue may have influenced the consistency of which indicators were rated to describe the contributions. It would not be prudent to over generalize these results due to the delimitation that the study focused only on one masters-level discussion forum in a multimedia systems course.

**Pilot Study**

The pilot study entailed having a group of 27 raters use the 19-indicator-point preliminary SCAFFOLD instrument to rate 15 discussion forum contributions. Although the raters were doctoral candidates, the raters were somewhat representative of college faculty. All raters had at least a master’s degree, and 25% had taught or were currently teaching online courses at their institution.

The threaded discussion forum transcript reviewed in the pilot study was taken from a master’s course in multimedia systems offered in the computer information systems curriculum. Although the raters did not necessarily have either education or experience in multimedia, the nature of the topic – What factors separate good (effective, productive) use of multimedia from bad (ineffective, or counterproductive) use of multimedia? – was general enough that all raters felt comfortable reviewing the discussion. The raters were given explicit instructions to rate the indicator as a descriptor. The quality of the content of the contributions was not evaluated.

**Data Analysis**

The 15 contributions assessed by the 27 raters offered a potential of 420 instances of using the 19-point preliminary SCAFFOLD instrument. In the analysis, a data reduction utilizing a principal component analysis was run on these data to determine if meaningful factors could be identified, affording the opportunity of condensing the SCAFFOLD instrument to a more manageable size.

**Results**

The data reduction resulted in identifying three components that accounted for 60% of the variance among the 19 indicators included in the preliminary version of SCAFFOLD. Table 3 presents the those components with the correlation coefficients for all the indicators included in each component. The three components identified in the data reduction appear to track closely to the three levels of reflection included in Mezirow’s [16] classification – content, process, and premise. Content level reflection entails acquiring facts and building skills. The indicators included in component 1 describe the contribution at the content level. Process level reflection entails developing problem solving ability; the indicators included in component 3 describe the contribution at the process level. Premise level reflection, the most cognitively demanding learning, entails an analysis and evaluation of the value and relevance of the subject matter. The indicators included in component 2 seem to describe the premise-level attributes of the contribution. Based upon this analysis, Table 4 presents an abbreviated version of SCAFFOLD.
The contribution was on topic of the following characteristics:

- Broadened: increased the scope of the discussion
- Originality: contained new ideas or approaches to the topic
- Problem: identified a worthy problem related to the topic
- Questioning: raised thoughtful questions about the topic
- Comprehensive: was complete, but not overly lengthy
- Evaluative: was evaluative, assessing the meaningfulness or validity of ideas being shared
- Synthesis: contained well formed, clear, connected, and synthesized ideas

Component 2

- Closure: helped lead to a conclusion on a topic
- Resolution: promoted cooperation to resolve issues of debate or disagreement
- Solutions: suggested meaningful solutions
- Summarizing: summarized the topic discussion overall
- Analysis: provided analysis of the problem being discussed

Component 3

- Acknowledging: responded to another contribution
- Clarification: supplied or sought clarification as needed in responses
- Reflective: interjected personal commentary or experiences
- Social: conversational or social in nature
- Resources: exchanged useful resources with others such as links or citations

Unfactedored*

- Error Free: contained accurate information
- Topical: was on topic

*Did not correlate with the other indicators in any component at a statistically significant level

### TABLE III

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadened</td>
<td>.738</td>
</tr>
<tr>
<td>Originality</td>
<td>.851</td>
</tr>
<tr>
<td>Problem</td>
<td>.957</td>
</tr>
<tr>
<td>Questioning</td>
<td>.958</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>.410</td>
</tr>
<tr>
<td>Evaluative</td>
<td>.442</td>
</tr>
<tr>
<td>Synthesis</td>
<td>.484</td>
</tr>
</tbody>
</table>

### TABLE IV

The contribution demonstrated content level reflection by including any of the following characteristics:

- Broadening the discussion
- Presenting new ideas
- Raising meaningful questions
- Making a comprehensive posting
- Assessing others contributions
- Presenting ideas in a clear, connected fashion

The contribution demonstrated process level reflection by including any of the following characteristics:

- Responding directly to postings of another
- Offering clarification of postings
- Interjecting personal commentary
- Being conversational
- Did not merely list resources such as links or citations

The contribution demonstrated premise level reflection by including any of the following characteristics:

- Bringing closure to the topic
- Promoting cooperation in resolving issues
- Suggesting meaningful solutions
- Summarizing the discussion
- Analyzing the problem being discussed

The contribution was error free and contained accurate information

The contribution was on topic

### REFERENCES


