THE METHOD (AND MADNESS) OF EVALUATING ONLINE DISCUSSIONS

Katrina A. Meyer, Ph.D.
The University of Memphis

ABSTRACT

In addressing how to evaluate online discussions, this paper will describe several concepts, tools, or frameworks that have been used in evaluations and discuss differences in approach based on instructor purpose, be it research, assessment, or learning. Then several common problems are described, including use of content analysis, identification of latent content, inadequate training of coders, lack of reliability, and choosing the correct unit of analysis. Two examples are provided of coding decisions made on portions of student discussions; these examples use two different frameworks to elucidate the process and its difficulties. Conclusions focus on the importance of following standard good research or assessment practice and preparing for a time-consuming and often frustrating coding process.

I. INTRODUCTION

Evaluating online discussions is neither as simple nor as straightforward as one might suppose; it involves answering important questions about the instructor's purpose, the student learning to be measured, and the application of coding procedure. With a focus on the instructor who is new to such evaluations, this paper is an exploration and explanation of what is involved in evaluating online discussions. It will attempt to provide answers to questions a novice instructor might ask when thinking about how to evaluate an online discussion held during a course.

II. THE INSTRUCTOR'S PURPOSE: A REVIEW OF THE LITERATURE

Every instructor has a purpose (and perhaps several) when deciding to evaluate the transcripts of online discussions. Three purposes seem to predominate in the published literature: to conduct research, to assess student learning, and to develop professional expertise. These are not incompatible aims, and all could certainly be pursued together or sequentially. However, it is more likely that one aim will predominate, and choosing one aim over the other may motivate a different approach and dictate a different methodology.

A. Research on Online Learning

Research on online discussions has focused on either the development or the testing of theory, and there are several theories to test. Moore's theory of *transactional distance* asserts that the distance between educator and learner is not solely a function of geographical distance but the result of increasing or decreasing dialogue (the transaction in transactional distance) between and among instructors, students, and content [1, 2]. Garrison et al.'s *community of inquiry* theory [3] posits the importance of three presences—cognitive, social, and teaching—that come together to create the online educational

experience. As distance education moves beyond a focus on structural and organizational matters, *transactional theory* [4], which focuses on collaborative and adaptable learning rather than purely independent or instructor-determined models, may become increasingly powerful for explaining the effectiveness of distance and online education. Drawing on constructivist learning theory, the theory of the *social construction of knowledge* [5] emphasizes the importance of social interactions and the social context in education, arguing that prior beliefs are challenged and new knowledge is constructed through discussion and sharing of alternative viewpoints. The use of *computer-supported cooperative work* [6] is another theoretical approach that emphasizes the social context—specifically, the use of computers to connect students and provide them with tools for interaction in order to develop their cooperative learning and work skills. The *communicative theory of collaborative learning* [7] focuses on language acts that are constituents of collaborative learning along two dimensions: learner orientation (learning, achieving ends, self-representation, and promotion) and knowledge domain (subject matter; norms and rules; and personal experiences, desires, and feelings). The *social interdependence theory of cooperative learning* [8, 9] depends on structuring the learning environment so that cooperation is not only possible but rewarded—in other words, students reach their own goals only when others in the class reach theirs. The theory of *community development* in an online discussion [10] has proposed three stages: from making friends, to getting involved, to experiencing camaraderie in the class. All of these theories are particularly useful for research on online coursework and, more particularly, online discussions.

Additional theories need to be developed in order to explain more specific phenomena: for instance, the precise ways in which online discussions are different from face-to-face discussions as well as unique characteristics of the online discussion as it occurs in current course management systems. And these theories—as well as the theories mentioned above—will need to continue to be tested by others.

Furthermore, theories about such concepts as community, communication, group work, interaction, and critical thinking developed for the pre-Internet world of classrooms and traditional pedagogies (such as lecture, small group work, and Socratic questioning) require testing in the online setting. In fact, theories developed to explain traditional instruction or classroom behavior too numerous to mention are probably ripe for application to online discussion.

Conducting research on the online classroom or online discussions requires the same commitment to careful design, appropriate measures, and thoughtful analysis as any research study. The results must be reliable, which can be a function of both the tool's design and its underlying theory or conceptual framework. It is wise to investigate the developer and his or her work in the literature, test the tool yourself, and assess the theory or concepts that support use of the tool. Unfortunately, the reliability of results is more likely to be a function of how instructors decide to use the tool in evaluating online discussions. In other words, the analysis might be flawed by the instructor's lack of understanding of the tool, not the use of multiple or poorly trained raters. It is easier for the instructor to do the rating of the discussions using the tool, but then the analysis can be seriously confounded by the instructor's preconceptions about the students and their abilities. This issue is an important one: for research purposes, solid rating practices and appropriate application of tools are essential.

B. Assessment of Student Learning

1. What to Assess

The assessment of student learning is an aim of many journal articles. For this purpose, there are many tools to choose from, but additional tools will likely be developed. This section provides a more detailed review of concepts or tools that attempt to capture different aspects of student learning.

Several articles have focused on the creation of *community* in online settings and the assessment of how and to what extent community has been developed. One theory of community building [10] stresses such evidence of community as the making of friends or acquaintances, involvement in long online discussions, and development of camaraderie. A rubric to assess community [11] focused on each student's participation in terms of its relevance, clarity, and contribution to community. The ability of online students to support other students may be another critical element of forming both community and the social construction of knowledge. For this purpose, the Transcript Analysis Tool (TAT) has been used to identify the most common support strategies as referential statements (statements that refer to others' previous comments), signatures, greetings, and horizontal questions (open-ended questions) [12].

Evidence of *collaboration* is important to the social construction of knowledge and is therefore important for student learning. Campos et al. [13, p. 44] characterized collaboration in terms of three levels—vague, modest, and strong—and several specific pedagogical choices, from "network-enhanced lectures" to a "collaborative learning project." Another coding scheme used the communicative theory of collaborative learning [7] to determine whether and how collaboration occurred.

Certainly *interaction*, which has been the focus of several published studies, is critical to collaborating, forming community, and constructing knowledge. Tools or concepts that tease out qualities of interaction are particularly useful, but, not surprisingly, different authors have assessed interaction differently. A discussion analysis tool [14, p. 30] identified patterns in interactions and helped determine which patterns promoted critical thinking; in this analysis, interactions were characterized by agreement or disagreement, arguments, personal experience, literature, formal data, personal or hypothetical actions or choices, evaluation or critiquing of arguments, summary, negotiation or conclusions, and process statements. Interaction has also been defined by specific types of contributions, such as participating, responding to others, providing feedback, and staying on topic [15]. Using a database that contained every sentence of a discussion, Hillman [16] coded each sentence by level or sentence purpose (organizing, lecturing, eliciting, responding, humanizing, idling, or not clear), by sentence mechanism (fact-stating, explaining, opining, performing, repeating, rating, rhetorical device, filler, and not clear), and by sentence content (person, action, procedure, content, supplies, and not clear). A software database system would be essential for such a multilayered analysis. Another tool, the previously mentioned TAT, allows instructors to categorize interaction in online discussions by designating each sentence as a question, statement, reflection, scaffolding, or quotation/citation [17]. The TAT can be helpful in analyzing how an online discussion is progressing and at what level an individual student may be operating. Another approach to assessing interaction is a rubric [18] used to gauge levels of disagreement, argument structure, level or type of interaction, and quality concepts that focus on student learning. Level, amount, and kind of interaction was also crucial to students' progress from feeling like outsiders to feeling like insiders [19], which was a personal and subjective assessment made by students through interviews that were subsequently analyzed through content analysis. One can see from this modest review of assessment practices that a number of approaches or tools are available to assess interaction, ranging from the complicated to the more intuitive.

While these studies have focused on assessing the individual student's learning, others have focused on the conversation or *discussion as a whole*. An approach to analyzing conversations as a whole is the conversation system model [20] that combines speech acts (assertion, promise, question, or command) and locus of conversation (first, second, or third person). This system leads to taxonomies that may be helpful in studying higher-order learning and conversations imbedded in different sociopolitical contexts. In a different assessment of discussion as a whole, individual student postings to an online discussion were graphed [21], showing modest support for the assertion that the group rose through higher levels of thinking in some, although not all, discussions.

Essential to judging whether a student is critically *engaged* in his or her own learning is a student engagement rubric [22] based on moving students beyond passive listening to a personal and deep engagement in their own learning. Cultivating engagement is essential for developing critical thinking and other skills in students, and students' development of critical thinking skills has been an important focus of several studies. The concept of *cognitive presence* [3] describes the performance of online students who are working through four steps in the discussion: triggering, exploring, integrating, and resolving a problem. In [6] critical thinking was characterized as a five-stage process—problem identification, definition, exploration, evaluation/applicability, and integration—and transcripts of online discussions were evaluated for 54 indicators of each stage. While this is certainly a thorough approach, instructors might find that applying 54 codes is a bit demanding.

Additional concepts from Garrison et al.'s [3] community of inquiry theory are *social presence* and *teaching presence*. Evaluating social presence—or the ability to project a real sense of self through online text—has been the focus of several studies [23, 24, 25, 26, 27] that use specific forms of evidence to identify this presence, including affective (e.g., expressing emotions, using humor), interactive (e.g., quoting another person's posting, asking a question), and cohesive (e.g., addressing the group as a whole) [27]. An alternative definition of social presence in [28] includes social context (including task, topics, and relationships), online communication (ability to type, read, and operate computers), and interactivity (including activities and communication styles). Evidence for teaching presence is based on several studies [29, 30] and can be found as acts associated with instructional design and organization (e.g., designing activities, using media effectively), facilitating discourse (e.g., identifying areas of disagreement, seeking consensus, recognizing contributions), and providing direct instruction (e.g., presenting content or questions, summarizing, diagnosing misconceptions) [29]. Perhaps a close analogue of teaching presence is *instructor immediacy* [31], which includes such evidence as instructor use of humor and student names, encouragement of discussion, following up on student comments, encouragement of future contact with students, and sharing of personal examples.

Another type of assessment examines *deep and surface learning* [32]. This approach uses content analysis to look for evidence of deep learning (e.g., looking for meaning, relating ideas, using evidence and logic, intrinsic motivation) versus surface learning (e.g., reproducing, staying inside course boundaries, unthinking, fear of failure, extrinsic motivation). This article provides an excellent example of how existing theories can be adapted to develop a new framework for analyzing online learning.

Other assessment tools have focused on individual students' *developmental or thinking skills*. Online discussions have been evaluated for evidence of students' development of ethical reasoning or reflective judgment [33] using Perry's schema for ethical and intellectual development [34] and King and Kitchener's [35] levels of reflective judgment. Bloom's taxonomy of educational objectives [36, 37, 38] provides the instructor with easy-to-use categories that capture levels of thinking, from knowing, understanding, applying, analyzing, evaluating, to creating.

Instructors interested in assessing their students' contributions to online discussions also need to be aware of research that focuses on individual and group differences based on whether the student is an introvert [39] or visual learner or active poster [40]. Evidence for gender differences [41] and changes that occur over time [42] is less conclusive, but the instructor using any assessment scheme ought to be aware that personality type, learning style, and gender may or may not explain some of the differences in student performance.

2. How to Assess

How are these analyses performed? Three approaches seem to be the most common. First, many of these studies use *content analysis* of the printed transcripts, looking for occurrences of certain words, phrases, ideas, concepts, or theoretical constructs. Examples of this approach are fundamental to using the community of inquiry theory [3, 27] and the TAT tool [17, 12]. As a result of these studies, we can more confidently claim that online students are learning, that this learning is of a certain quality or type (e.g., critical thinking, problem solving), and that instructors can introduce online discussions in their distance education or on-campus classes with some confidence that students will benefit from the experience. However, content analysis presents certain challenges to reliability, which will be discussed in a later section.

Second, *rubrics* allow an evaluator to assess the extent to which a quality is present along a continuum, and rubrics exist to assess interaction [43, 44], engagement [22], and critical thinking [45]. These tools require training and careful analysis of the discussions so that fine distinctions along the continuum can be made fairly and reliably.

Third, a number of *frameworks*—or approaches to classifying, defining, or critiquing student contributions—also exist, such as those mentioned in the previous section: Bloom's taxonomy of educational objectives [38], Perry's multilevel schema of ethical and intellectual development [33], and King and Kitchener's approach to identifying students' different levels of reflective judgment [35]. While these frameworks date from the pre-Internet era, they are valuable for assessing the level at which students are operating and have been applied to online discussions [33]. New frameworks will undoubtedly be developed specifically for online discussions. Therefore, it is necessary to stay familiar with the current research to see how useful some approaches may be.

How do you select the best tool for your purpose? The first step is to know the purpose of the evaluation—are you evaluating interaction, critical thinking, or the acquisition of particular content? Are you evaluating the discussion as a whole, the contributions of individual students, or your own contributions? Guidelines for choosing or designing assessment tools for online discussions are the same as those for assessing regular classwork [46], and students should be informed about such guidelines beforehand: that is, have clear objectives for the discussion and be clear with students about what will be evaluated and how. The second step is to become familiar with the range and variety of tools available. No one tool is appropriate for all purposes, and staying informed about current tools (and ones that are being developed) is the best way to ensure that you aren't fulfilling Maslow's caution: "To the man who only has a hammer in the toolkit, every problem looks like a nail."

C. The Instructor's Learning

But perhaps the initial purpose of evaluating online discussions is one of gaining professional expertise: instructors need to learn to use a new tool in order to ensure that their efforts will result in useful and/or reliable results should they assess student learning or conduct research. Since some tools are more intuitive or self-explanatory than others, preparation is necessary. Whenever scholars have developed a body of research to support or explain a particular tool or framework, instructors would be well advised to become familiar with the books or journal articles that describe and test its assumptions, critical issues, and significant findings before proceeding. This review, although time-consuming, might elucidate problems that would preclude the tool's use in an online setting or uncover misunderstandings about the tool that the faculty person may hold.

Instructors need to explore whether the tool is easy and usable, which requires an assessment of the clarity of the distinctions among levels or concerns. Some frameworks require extensive preparation and training, such as Perry's developmental schema [34]. Although well-read on the subject, the author was often hard pressed to determine differences in students' postings to an online discussion [33]. Distinctions made along a continuum, as is the case with rubrics, may also require some preparation and training and would lead to inconsistent findings or interpretations if undertaken by a poorly prepared instructor.

D. Implications for Different Purposes

Let us return to the decision about instructor purpose: is it to test theory or assess student learning or teach oneself a new skill? In each case the answer might involve different choices. The testing of theory might entail the application of more rigorous measures, including opting for independent raters who are neither the instructor nor the researcher, and calculating and reporting interrater reliabilities. Issues pertaining to experimental design should be carefully considered—for example, controlling for student differences if students cannot be randomly assigned to different treatments, and eliminating bias in the design of the instructional task so that results will not be unfairly affected.

Assessing student learning might argue for a more tailored approach that takes into account a number of issues. For example, what is the instructor's purpose for assessment: is it to grade students, improve the course, or prove to accreditors that learning occurred? Also, the amount of time available to conduct the assessment is important, because with some tools the process can be completed in days while other tools require more time. While multiple raters may not be needed, the instructor still needs to have confidence that his or her rating of the discussions is accurate and reliable, which means that he or she must understand the tool thoroughly, avoid rating when tired, and perhaps assess discussions—or postings to the discussion—without student names in order to avoid prejudgment.

Or is it more important for the instructor to improve his or her own understanding of an assessment tool, online discussions, and student learning, rather than to produce replicable results? In this case, perhaps some of the restrictions of traditional research design can be relaxed, and the reflection and learning of the faculty person becomes paramount.

It is important to remember that while each approach has value, each purpose may require different designs, different methods, and different standards for publication. If the instructor is a novice, a first study might involve understanding and developing expertise with coding discussions. The instructor's learning is predominant, and any article resulting from such an undertaking will likely be evaluated by standards of reflective learning, including recursive analysis of the "data" or insights, citation of disconfirming evidence or exceptions to an emerging understanding, and maintenance of the same standards applied to qualitative or ethnographic research [47]. This may seem excessive, but if the instructor's learning is intended for publication, readers need to know that the learning described is as fair and thoughtful as possible and not somehow flawed.

Or perhaps the evaluation is an attempt to assess student learning, which requires that codes be fairly and reliably determined. An article resulting from such work might illuminate what was learned and changed as a result of the assessment, but it would also necessarily provide sufficient detail about the context of the discussion and the method of coding so that readers can have confidence in the instructor's work and interpretation. Lastly, for research studies, the same requirements for publication in research journals would probably apply, including a full and complete explanation of the theory under investigation or the theory development process, the study methodology and analyses, and the findings and interpretations in light of the previous literature and theory.

III. TECHNICAL ISSUES IN THE EVALUATION OF ONLINE DISCUSSION

There are at least four technical issues that appear repeatedly in the research studies and articles on online discussions, and each deserves serious consideration. They relate to using content analysis, deciding upon the criteria for handling latent meaning when coding, determining the unit of analysis, and using multiple raters. For those who may be interested in additional methodological issues, please see Rourke et al. [48] for an excellent detailed discussion.

A. Defining the Issues

Content analysis, a standard research technique in qualitative research designs, involves the identification of themes across multiple interviews or documents, even though different terms for similar concepts may be used. It allows the researcher to note different perspectives or conceptualizations of the chosen theme and how individuals characterize or approach the topics under discussion. Content analysis may be objective when it involves counting the number of times a particular word or set of words is used (this is also known as manifest content) or it may be subjective, depending upon the rater's interpretation of the meaning of what has been written (this is latent content, which is discussed next). However, "an excessive degree of subjectivity should signal to the research team that further refinement is needed in category definition or coding protocol" [48]. In other words, subjectivity may mean the tool is inappropriate for the task at hand, the tool is too unspecific, or the raters need more training. This discussion of subjectivity may not be important to the individual instructor who wishes only to assess a particular quality of his or her students or an aspect of their learning. But if the ultimate goal is to produce replicable and reliable research to guide others in their instructional choices, then one corollary aim should be to lessen subjectivity in the assignment, coding, and interpretation of content analyses of online discussions.

Subjectivity or *latent content* in students' postings to an online discussion become a problem when coders' interpretations of meaning become variable, personal, or idiosyncratic. There are several ways of handling latent content, including only coding obvious examples with clear meaning that fit a certain classification and ignoring the "shades of gray" [6] or delaying analysis of latent material until the analysis and interpretation stage [49]. Depending on how stringent the researcher wishes to be, latent content can be determined by the coders' gut feeling, which may provide useful information to those wishing to assess student learning but is not suitable for research intended to test hypotheses based on a formal theory or theories.

The *unit of analysis*, which can also vary, deserves more attention by the researcher than may appear to be necessary. The unit of analysis may be a particular word, phrase, sentence, paragraph, message, or single posting, or units defined by a theme or purpose. Choosing a unit of analysis affects the subjectivity and reliability of the codings, the likelihood of dealing with more instances of latent content, the difficulty or ease of the coding, and, finally, the meaning of the analysis. Some instructors may be concerned about a student's understanding of a particular term, his or her expression of a complete thought or sentence, the presentation of an argument, or the interactions among students. Therefore, it is important to let the goals of the discussion and purposes of the assessment guide this choice.

Many research studies do not use *multiple raters* to code the content in online discussions. This may be owing to a number of factors, including the instructor's preference for working alone or a lack of interested colleagues to help with coding. Researchers may not have the time to train other coders or the money to pay them, or perhaps the aim is simply to collect data about the learning of a given set of students rather than to produce reliable findings. In other words, there may be understandable reasons for not using multiple raters, despite the greater reliability that might result from their use. But as the

motivations to produce more rigorous and replicable results increase, the use of multiple raters should also increase.

These issues argue for the novice evaluator of online discussions to become conversant with how to prepare a reliable content analysis, detect and handle latent meanings, and distinguish among multiple interpretations. Instructors should fully train multiple raters in the analytical tool and use them when the purpose of the exercise is to produce replicable research.

B. Coding Quandaries and Examples

This section provides two examples of the coding of online discussions that occurred in a doctoral-level educational leadership class on the financing of higher education. The ten students were given one week to answer each question, and their contributions were used as a course assessment—that is, points were assessed both for their involvement and the quality of their contributions. Quality was defined in terms of answering the question with concrete suggestions or new ideas. This was a novel use of the online medium for the group as well as the instructor, although the group was experienced with online discussions, having used them in several classes, and ready to try something different. The instructor wished to find out whether the course assessment—a modified version of a final take-home exam—could be carried out as an online discussion and whether it would be a valuable and accurate assessment. The students' contributions would be public and therefore known to everyone in the class, rather than to the instructor only as in a traditional exam. The instructor also wished to determine whether the medium could provide a test of student knowledge without generating the stress of test taking, since students could log on to the discussion anytime in the course of the week, contribute from their home or office computers, and return to add new ideas generated from their readings or other students' postings.

1. The Discussions and Coding Schemes

The postings in Table 1 are in response to the question "How do we retrench, cut budgets, and make progress?" This question sought to pull together approaches to cutting college and university budgets so that an institution can continue to achieve its objectives. Postings are not presented in the order they occurred, and all names of students and institutions have been replaced, although references to specific articles have been left in. The framework used to code these contributions is based on Garrison et al. [3], who proposed a four-stage process of critical thinking: (1) triggering, or posing the problem; (2) exploration, or the search for information; (3) integration, or construction of possible solution; and (4) resolution, or critical assessment of solution. The possible coding—or Garrison et al. stage—is offered in the second column, along with a short justification for this particular choice based on the sometimes much longer posting. This framework has been used by the author [33] to analyze online discussions. Student references to the "Chronicle" are to *The Chronicle of Higher Education*, a national weekly journal for higher education administrators and faculty (http://www.chronicle.com).

Table 1. Evaluation of Discussion on Budget Cutting Using the Garrison et al. Critical-Thinking Framework

Posting	Possible Coding
How do we retrench, cut budgets, and make progress?	*1. Triggering Question*
I was going to make some smart comment about if you figure it out (how to retrench, cut budgets and make progress), let me know. We certainly can use some help here at Institution A.	*2. Exploration* In this posting, the student is searching for information from other students. The posting continues with several suggestions for cutting budgets, but this portion of the posting more nearly

	approximates the spirit of exploration.
What if we focus on strengthening our programmatic niches? We can't be all things to all people. Or make sure that new initiatives fit within the institution's mission. Or use technology for administrative activities … Or form partnerships with other institutions to share courses.	*2. Exploration* This posting is composed of a brief exploration of several suggestions that are not brought together or integrated.
Here are some things happening at Institution B to address this [remainder of posting is a list of 10 actions institution is pursuing].	*2. Exploration* This posting is composed of a list of several suggestions that are not brought together or integrated.
I just read an article in this week's Chronicle about Reclaiming Fridays … The article talks about how classrooms (and campuses) empty out on Fridays. So the idea is to offer more classes on Fridays. But then there is the problem of getting students to attend on Fridays (usually, there is about a 50% reduction in attendance for Friday classes) even when they are scheduled. How cost effective is it to have empty classrooms? … we could encourage faculty to have exams, quizzes, projects due on Friday … right now the week is getting shorter and the tuition is getting higher.	*3. Integration* This is an example of constructing a solution to the problem of using resources more wisely.
Higher education has to do strategic planning and decide what we can and cannot do. We can't be all things to all students.	*2. Exploration* This brief suggestion is more of an exploration than a solution resulting from a thoughtful integration of ideas.

We've just eliminated several assistant deans at Institution A and instead have chairs for five divisions rather than 26 … will this save money in the long run or will it streamline effort … or both?	*2. Exploration* This posting is an excellent example of exploring an idea and finishing with a possible triggering question that did not trigger further discussion.
One strategy Institution E [a two-year technical college] used was to partner with four-year institutions to provide general education courses. If colleges could evaluate their course offerings and partner with other institutions to share the cost of low-volume courses, they could reduce costs. Our program avoided the chopping block by partnering with other institutions … we need to look at a more standardized method to determine costs and work backwards from that to determine where cuts can be made.	*4. Resolution* This long posting starts with a simple suggestion (sharing courses), proceeds through an extensive description of how a degree program was saved by several actions, and finishes the description with a new approach that may form a resolution to the problem of cutting costs.
The words "sharing courses" got my attention. That is something we are trying to do here ... But until we actually cut programs, I don't see this as a money saver.	*3. Integration* This posting included a detailed description of what was happening to "share" courses, which led to the final comment and evaluation of the approach.

One small thing that is happening at Institution B is the full-year utilization of the education building, including Saturday and Sundays. We have Weekend College for working people and many summer activities ... the building is never quiet!	*3. Integration* This is an example of constructing a solution to the problem posed to the group.
There are many areas under scrutiny at Institution C. Our new dean has declared that classes with fewer than 6 students will no longer be allowed ... this makes sense to me but we are a small liberal arts college and it may hurt some programs.	*4. Resolution* This is a portion of a much longer posting that provides a critical assessment of the proposal to save resources by cutting classes.

In Table 2, the discussion was in response to the question "Where shall new sources of funding be found?" The question was intended to generate ideas for revenue sources for colleges and universities aside from the raising of tuition, and the postings are evaluated using Bloom's taxonomy [36,37] as revised by Anderson and Krathwohl [38]. The Bloom taxonomy benefits from being a foundational tool well known to individuals trained in the field of education. It ranks educational objectives into six levels from lower to higher: Know, Understand, Apply, Analyze, Evaluate, and Create. Perhaps due to the discussion's occurring at the end of the course and among doctoral-level students, there were few examples of postings at the lower Bloom levels.

Table 2. Evaluation of Discussion on New Revenue Sources Using Bloom's Taxonomy

The Chronicle had an article on funds ... where universities are teaming up with big-ticket investment companies ... Harvard [is] managing the funds and doing quite well attracting investors. The article stated, "there is a cachet for some donors to have their money managed by the Harvard Management Company." Does anyone have any experience with this type of thing? Is it a good idea?	*Understand* This student is not attempting to evaluate or analyze this proposal reported in the Chronicle but appears to be trying to understand the idea and asking for help from the group to do so.
I saw a Chronicle article about colleges allowing limited advertising ... "it is a tough way to generate revenue" ... so, obviously this isn't an immediate viable method of fundraising. However, I wonder how long it will be before we see more of this? A rich and powerful alumnus ... roadside signage ... adopting of freshmen ...	*Apply* In this posting, the student allows the article quoted to do the evaluation of the proposal but continues with applying or exploring several examples of what might be advertising opportunities, without evaluating examples.
One new source of funding as we move into a global market will be overseas institutions of higher education, government agencies, technology companies, and others that we can form partnerships with. [Student gives specific example from Institution D.]	*Apply* The student applies this new funding source by describing current efforts at Institution D to pursue overseas partnerships.
I know our continuing education and workforce training groups are working on providing more workforce training to generate revenue. And we could provide workshops for area businesses or career training (like conflict resolution and stress	*Apply* This is a listing of possible revenue sources without much analysis or evaluation.

management). [Article] proposes colleges should generate revenue by providing corporate services.	
Corporate funding has been around for some time in one form or another. A lot of research funds come from business or business foundations. It still makes me very uncomfortable. There seems to be a fine line between working with a corporation and working for a corporation. The same goes for government. There are always strings attached.	*Evaluate* One could argue that this posting is both analysis and evaluation, although there appears to be a stronger sense of judging value in the tone of the posting than a dispassionate analysis of the pros and cons of seeking corporate funding.
I just read an angry article about "retrenchment by attrition"—the loss of full-time faculty positions following retirements or resignations. Faculty are angry because budget cuts have required the institution hire younger faculty who are cheaper or not filling the position. But while everyone knows we spend too much, it is always someone else's program that should be cut.	*Analyze* This is primarily an analysis of faculty responses and responsibilities toward budget issues, although at the end the student offers an evaluation or interpretation of usual faculty responses.

I have no great ideas for new sources of funding. Institutions seem to be picking at the same old bones: students (tuition & fees), government (taxpayers), alumni, and business. It seems to me a better idea to offer quality services and programs that have value and serve a niche than to think about funding sources first.	*Analyze* This student is proposing an alternative solution, which might be a "create" response were it an original proposal and not derived from reading the literature.
I had a conversation with a dean who told me there was too much bureaucracy on campus (which was preventing him from doing what he wanted) AND then said my office should be doing more to ensure X, Y, and Z don't happen. I had to ask, "How can I do that without adding rules, regulations, or procedures?"	*Evaluate* The posting presents an unwinnable conundrum for this administrator and represents an analysis of the dean's contrary suggestions that has a strong evaluation tone.
What about "marketing" the good speakers on a faculty? Every institution has people with charisma and who are recognized through their books and publications ... it just seems they might have something to offer.	*Create* This may be an example of the student creating a new revenue source (it does not appear tied to an idea garnered while reading the literature).
The Chronicle this morning mentioned a new fund-raising idea ... [an institution] is initiating a charge of $100 per intercollegiate team that each student participates in over the year. ... The monies generated go back to athletics, [which frees up] institutional resources.	*Analyze* This posting reports on a new fee and then analyzes how it could help an institution. The posting is free of judgmental language, as might be more common in a posting that evaluates.

2. Problems and Challenges in Online Discussion Coding

The comments or justifications for the decisions made in the second columns in Tables 1 and 2 begin to illustrate several problems experienced when coding discussions. First, what is difficult to make visible is the recursive nature of the process—the coder's repeated reference to the coding schema to reread the

definitions of each code. The last posting in Table 2 that begins "The Chronicle this morning…" provides an example: Is this analysis or an application of an idea found in a news source (i.e., The Chronicle)? Is it evaluation? The instructor needs to reason whether it is Apply, Analyze, or Evaluate. The first reading does not lead to an immediate decision, so a second, third, or additional readings are required. The response seems to go further than just Report or Apply; it develops a use for the funds that was not in the original article, but it does not seem to evaluate the idea (the terms used by the student do not seem particularly value laden) beyond the fact that it generates funds. This is an example of how coding takes time and sometimes requires that the evaluator decide among several codes to determine the best or most appropriate one. The instructor, who is familiar with the articles and course content, may be the best person to make such decisions, but that does not mean that another person well versed in the same material could not also do the coding.

Second, the coder must often work to unravel the nature of the posting by distinguishing between two codes on the way to a final decision. The comments in the paragraph above and several entries in the second column provide a sense of this process. While understanding the tool is crucial, understanding the students' postings may also require examining the intentions of the student, the level of thought, the words used, the construction of the sentence, the context of the sentence within the entire posting, or the posting within the context of the entire discussion. This not only takes time but is taxing on the coder.

Third, the unit of analysis will affect your process and results. In Tables 1 and 2, the postings in the first columns are much longer than the abstracts in the second columns, and students often touch upon many topics and switch approaches to the material. If the unit of analysis is an entire posting, a decision must be made whether what is important is the predominant nature of the posting (for example, is it more evaluative than analytical) or the final level of thinking reached by the end of the posting. For more verbose students, this might argue for using a smaller, more discrete unit of analysis, like a sentence or paragraph. But for faculty interested in the student's reasoning process, perhaps the level of reasoning reached by the end of the posting is of greatest interest and captures the end result of that student's thinking.

Lastly, the process is time-consuming and requires serious commitment on the part of the coder. Frequent breaks are necessary, as are frequent returns to the code or tool. The time and frustrations involved can be maddening, and one can soon come to doubt one's sanity.

IV. RECOMMENDATIONS

Teachers and researchers now have excellent opportunities to use the printouts of online discussions for more in-depth evaluation of students and their learning, or they can use these same printouts to reflect upon the discussion process and outcomes in order to improve future discussions. However, the new opportunities presented by online education come with the requirement to retain good practices of research and assessment. Even though we may feel as if we are exploring new modes of teaching and creating new ways of learning, this feeling of experimentation should not extend to decisions about research design or assessment practice.

Good research requires careful design, understanding the theory and the research behind it, control of extraneous factors, a reliable and valid coding schema, resolution of issues of content analysis and latent content, and the training of multiple coders in the scheme. Good assessment requires that the purpose of the discussion is clear to students, the rubric or framework used to assess the discussion is valid, the coder is sufficiently trained in the rubric or framework, and the instructor can and will use the results for

improving subsequent discussions. And the instructor learning a new tool must commit to being (1) well-informed about the tool, the theory it is based on, the research behind it, and the advice others have written about it, and (2) as rigorous in the use of the tool as time and talent will allow.

Several theories are applicable to online discussions and beg for further research. And theories that may never have been applied to online learning—those developed and applied to traditional classroom activities—must be tested in the online setting to determine whether they are applicable to online learning or require modification. It is not clear whether online discussions are markedly different from face-to-face discussions, and much work must still be done to flesh out whether the two are different and, if so, in what respect. At least for now, the null hypothesis that should guide our investigations is that the two forms of discussions—online and face-to-face—are the same, and the same theories should apply to both modes.

Numerous tools are available for faculty interested in assessing various aspects of student learning: critical thinking, collaboration, and engagement, among several others. While several of these have been developed specifically for assessing online discussions, concepts developed for traditional, in-class discussions may still be tested in the online setting. As new aspects of student learning arise, new tools will be developed, tested, and shared so that others may assess their value. Because the purpose of assessment tools is to measure student learning, we may never exhaust new ways of finding proof that different aspects of learning have occurred.

If you are considering undertaking research or assessment of student online discussions, be prepared. It is difficult and frustrating, intellectually demanding and demanding of one's time. Nevertheless, additional evaluations of online discussions are needed: we need to know if our current theories of learning and discussion apply to learning and discussion online, whether our decisions as instructors and designers of online activities are having the result intended, and how we can make improvements in practice so that future online discussions are better than those that have gone before.

V. REFERENCES

1. **Moore, M. G.** Three types of interaction. In K. Harry, M. John, and D. Keegan (Eds.), *Distance Education: New Perspectives*, 19–24. New York: Routledge, 1993.
2. **Chen, Y. and F. K. Willits.** A path analysis of the concepts in Moore's theory of transactional distance in a videoconferencing learning environment. *Journal of Distance Education* 3(2): 1998.
3. **Garrison, D. R., T. Anderson and W. Archer.** Critical thinking, cognitive presence, and computer conferencing in distance education. *The American Journal of Distance Education* 15(1): 7–23, 2001.
4. **Garrison, R.** Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. *International Review of Research in Open and Distance Learning* 1(1): 2000.
5. **Gunawardena, C. N., C. A. Lowe and T. A. Anderson.** 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, 1997.
6. **Newman, D. R., B. Webb, and C. Cochrane.** A content analysis method to measure critical thinking in face-to-face and computer supported group learning. *Interpersonal Computing and Technology* 3(2): 56–77, 1995.
7. **Treleaven, L.** Evaluating a communicative model for web mediated collaborative learning and design. *Australian Journal of Educational Technology* 19(1): 100–117, 2003.
8. **Aviv, R.** Educational performance of ALN via content analysis. *Journal of Asynchronous Learning Networks* 4(2): 53–72, 2000.

9. **Aviv, R., Z. Erlich, G. Ravid and A. Geva.** Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Network*s 7(3): 1–23, 2003.
10. **Brown, R. E.** The process of community-building in distance learning classes. *Journal of Asynchronous Learning Networks* 5(2): 2001.
11. **Edelstein, S. and J. Edwards.** If you build it, they will come: Building learning communities through threaded discussions. *Online Journal of Distance Learning Administration* 5(1): 2002.
12. **Fahy, P. J.** Indicators of support in online interaction. *International Review of Research in Open and Distance Learning* 4(1): 2003.
13. **Campos, M., T. Laferrière and L. Harasim.** The post-secondary networked classroom: Renewal of teaching practices and social interaction. *Journal of Asynchronous Learning Networks* 5(2): 2001.
14. **Jeong, A. C.** The sequential analysis of group interaction and critical thinking in online threaded discussions. *The American Journal of Distance Education* 17(1): 25–43: 2003.
15. **Muirhead, B.** Enhancing social interaction in computer-mediated distance education. *Ed at a Distance* 15(40): 2001.
16. **Hillman, D. C. A.** A new method for analyzing patterns of interaction. *The American Journal of Distance Education* 13(2): 37–47, 1999.
17. **Poscente, K. R. and P. J. Fahy.** Investigating triggers in CMC text transcripts. *International Review of Research in Open and Distance Learning* 4(2): 2003.
18. **Spatariu, A., K. Hartley and L. Bendixen.** Defining and measuring quality in online discussions. *Journal of Interactive Online Learning* 2(4): 2004. http://www.ncolr.org/jiol/issues/PDF/2.4.2.pdf.
19. **Wegerif, R.** The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks* 2(1): 1998.
20. **Luppicini, R. J.** Toward a conversation system modeling research methodology for studying computer-mediated learning communities. *Journal of Distance Education* 17(2): 87–101, 2002.
21. **Meyer, K. A.** The ebb and flow of online discussions: What Bloom can tell us about our students' conversations. *Journal of Asynchronous Learning Networks* 9(1): 2005.
22. **Brown, G.** Guide to Rating Critical Engagement. Pullman, WA: Washington State University. Published online via subscription, September 2004.
23. **Gunawardena, C. N. and F. J. Zittle.** Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *The American Journal of Distance Education* 11(3): 6–26, 1997.
24. **Oren, A., D. Mioduser and R. Nachmias.** The development of social climate in virtual learning discussion groups. *International Review of Research in Open and Distance Learning* 3(1): 2002.
25. **Richardson, J. C. and K. Swan.** Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks* 7(1): 68–88, 2003.
26. **Rourke, L. and T. Anderson.** Exploring social interaction in computer conferencing. *Journal of Interactive Learning Research* 13(3): 257–273, 2002.
27. **Rourke, L., T. Anderson, D. R. Garrison and W. Archer.** Assessing social presence in asynchronous, text-based computer conferencing. *Journal of Distance Education* 14(3): 51–70, 1999. http://cade.athabascau.ca/vol14.2/rourke_et_al.html.
28. **Tu, C. and M. McIsaac.** The relationship of social presence and interaction in online classes. *The American Journal of Distance Education* 16(3): 131–150, 2002.
29. **Anderson, T., L. Rourke, D. R. Garrison and W. Archer.** Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks* 5(2): 2001.
30. **Shea, P. J., A. M. Pickett and W. E. Pelz.** A follow-up investigation of "teaching presence" in the SUNY Learning Network. *Journal of Asynchronous Learning Networks* 7(2): 61–80, 2003.
31. **Hutchins, H. M.** Instructional immediacy and the seven principles: Strategies for facilitating online courses. *The Online Journal of Distance Learning Administration* 6(3): 2003. http://www.westga.edu/~distance/ojdla/fall2003/hutchins63.html.
32. **Gerbic, P. and E. A. Stacey.** Purposive approach to content analysis: Designing analytical frameworks. *The Internet and Higher Education* 8(1): 45–59.

33. **Meyer, K. A.** Evaluating online discussions: Four different frames of analysis. *Journal of Asynchronous Learning Networks* 8(2): 2004.
34. **Perry, W. G.** *Forms of Ethical and Intellectual Development in the College Years: A Scheme.* San Francisco, CA: Jossey-Bass, 1999.
35. **King, P. M. and K. S. Kitchener.** *Developing Reflective Judgment.* San Francisco, CA: Jossey-Bass, 1994.
36. **Bloom, B. S. and D. R. Krathwohl.** *Taxonomy of Educational Objectives: The Classification of Educational Goals.* New York: Longmans, Green, 1956.
37. **Krumme, G.** Major Categories in the Taxonomy of Educational Objectives (Bloom 1956). http://faculty.washington.edu/krumme/guides/bloom.html.
38. **Anderson, L. W. and D. R. Krathwohl** (Eds.). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman, 2001.
39. **Carnevale, D.** Introverts do well in online chats. *The Chronicle of Higher Education* 50(16): A29, 2003.
40. **Rovai, A. P. and K. T. Barnum.** On-line course effectiveness: An analysis of student interactions and perceptions of learning. *Journal of Distance Education* 18(1): 57–73, 2003.
41. **Fahy, P. J.** Use of linguistic qualifiers and intensifiers in a computer conference. *The American Journal of Distance Education* 16(1): 5–22, 2002.
42. **De Simone, C., Y. Lou and R. F. Schmid.** Meaningful and interactive learning supported by the use of metaphor and synthesizing activities. *Journal of Distance Education* 16(1): 2001.
43. **Roblyer, M.D. and W. R. Wiencke.** Design and use of a rubric to assess and encourage interactive qualities in distance courses. *The American Journal of Distance Education* 17(2): 77–98, 2003.
44. **Roblyer, M. D. and L. Ekhaml.** How interactive are YOUR distance courses? A rubric for assessing interaction in distance learning. *The Online Journal of Distance Learning Administration* 3(2): 2000.
45. **WSU Critical Thinking Project.** The Critical Thinking Rubric. Pullman, WA: Washington State University. http://wsuctproject.wsu.edu/ctr.htm.
46. **Moore, W. S. and E. Rousso.** Principles for Designing and Evaluating the Quality of Online Assignments/Assessments. Unpublished document.
47. **Goetz, J. P. and M. D. LeCompte.** Ethnography and qualitative design in educational research. Orlando, FL: Academic Press, 1984.
48. **Rourke, L., T. Anderson, D. R. Garrison and W. Archer.** Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education* 12: 8–22, 2001. http://aied.inf.ed.ac.uk/members01/archive/vol_12/rourke/full.html.
49. **Holsti, O.** Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley, 1969.

VI. ABOUT THE AUTHOR

Katrina A. Meyer is associate professor of higher and adult education at The University of Memphis, specializing in online learning and higher education. She is the author of *Quality of Distance Education: Focus on On-Line Learning*, a 2002 publication of the ASHE-ERIC Higher Education Report Series. For over three years, she was director of distance learning and technology for the University and Community College System of Nevada. Before that, she served over eight years as associate director of academic affairs for the Higher Education Coordinating Board in the state of Washington and was responsible for technology planning and policy related to online learning.