Ask and you will receive: how question type influences quantity and quality of online discussions

Megan E. Bradley, Lindsay R. Thom, Jennifer Hayes and Cindy Hay

Megan Bradley is an associate professor of Psychology at Frostburg State University (FSU). Her research interests include online education, higher-order thinking and deep approaches to learning. Lindsay Thom, a graduate student and Jennifer Hayes an undergraduate student, are both research assistants. Cindy Hay, an adjunct professor at FSU, is obtaining her master’s degree in Distance Education at University of Maryland, University College. Address for correspondence: Dr Megan E. Bradley, 101 Braddock Road, Psychology Department, Frostburg State University, Frostburg, MD. 21532, USA. Email: mbradley@frostburg.edu.

Abstract
This study examined how question types influenced the quantity and quality of undergraduate students’ online submissions. Discussion questions were structured based on one of six question types: direct link, course link, brainstorm, limited focal, open focal and application. Transcripts of eight online discussions involving 114 participants were analysed on these dependent variables: word count, degree of answer completion, and level of higher-order thinking. Results indicated the question types differentially influenced students’ submissions on all dependent variables. Limited focal question type, followed by brainstorm, open focal and direct link types, was most influential for word count and degree of answer completion. Course link, brainstorm and direct link types were most influential in generating higher-order thinking, although students mainly engaged in lower-order thinking across all question types.

Introduction
In his seminal work on learning, Bloom (1956) recommended that educators promote complex thinking in students through test questions and learning activities. He outlined six hierarchical levels of thinking within the cognitive domain: knowledge, comprehension, application, analysis, synthesis and evaluation. Subsequent researchers have divided these six levels into the categories of ‘lower-order’ and ‘higher-order’ thinking. Although there is some disagreement about such a division and where the centre point should be (Krathwohl, 2002; Schrire, 2006), many researchers view the first three levels as constituting lower-order thinking and the last three levels as constituting higher-order thinking (Andrews, 1980; Gilbert & Dabbagh, 2005; Notar, Wilson & Montgomery, 2005; Schrire, 2006). One learning activity used to promote higher-order thinking is class discussion.
The theoretical underpinning of learning via class discussion is the social-constructivist perspective (Vygotsky, 1978), which posits that students actively create their knowledge through interactions with peers. Much of the research on the effectiveness of class discussion in face-to-face and online settings have focused on how both settings have context-specific advantages. Face-to-face discussions tend to have greater efficiency, immediacy of feedback, no technological issues, greater perceived interactivity and important verbal and non-verbal communication cues present (Tiene, 2000; Wang & Woo, 2007). Some advantages of asynchronous online discussions include convenience and flexibility of access, increased reflection time for posts, less intimidating context for introverted students and increased number of discussion comments (Christopher, Thomas & Tallent-Runnels, 2004; Meyer, 2003; Paskey, 2001; Walther, 1996).

Research on students’ higher-order thinking in class discussions has focused on graduate students’ online discussion comments, as analysed by various coding schemes that define higher-order thinking as involving critical thinking skills or Bloom’s Taxonomy. Using a critical thinking coding scheme, Garrison, Anderson and Archer (2001) found only 17% of comments reflected higher-level thinking. A follow-up study by Meyer (2003) using the same Garrison et al (2001) coding scheme found a greater percentage of graduate student submissions evidencing higher-level thinking; however, 69% of student comments were rated as lower-order. Using a coding scheme based on Bloom’s Taxonomy, Gilbert and Dabbagh (2005) found that 16–26% of online postings were rated as higher order. In contrast, Schrire (2006) found most students demonstrated higher-order thinking, as defined by levels of Bloom’s Taxonomy. The current study is one of the first to examine undergraduate students’ level of thinking via Bloom’s Taxonomy.

Current research explored ways to improve students’ online contributions and higher-order thinking through various manipulations such as course characteristics, moderating personnel and task complexity. Gilbert and Dabbagh (2005) demonstrated how including facilitator guidelines and scoring rubrics led to improvements in graduate students’ higher-order thinking in online posts. Rourke and Anderson (2002) found graduate students preferred peer moderators to an instructor moderator and that discussions facilitated self-reported higher-order learning. Finally, Schellens, Van Keer, Valcke and De Wever (2007) showed that while certain undergraduate student characteristics (e.g., attitude towards learning) contributed to levels of knowledge construction in online discussions, task complexity also mattered; specifically, students had higher levels of knowledge construction with moderately complex tasks compared to simple or overly complex tasks. Taken together, these studies indicate that quality of online discussion can be improved.

The present study focused on how the structure of discussion questions influenced undergraduate online discussion submissions. Very little research has been conducted in this area despite the importance of phrasing questions in an optimal manner. Indeed, a lack of preparation for a discussion could lead to questions that are poorly phrased, unclear, oscillating or too multifaceted (Andrews, 1980). One study on graduate students (Christopher et al, 2004) examined the level of higher-order thinking in student-
moderated prompts with the level of higher-order thinking in responses. No differences were found; however, no attempt was made, at the time prompts were created, to have them differ in the level of thinking nor was the structure of the questions associated with the prompts analysed. In the current study, questions were intentionally structured into one of six distinct types prior to the discussion to examine if responses would differ.

An earlier study by Andrews (1980) specifically examined how the structure of questions impacted student discussion in a face-to-face undergraduate course. By examining transcriptions of discussions, 10 discussion question types were found and the corresponding number of student responses per type was analysed. Although no formal statistics were conducted, Andrews stated that three of those question types were superior in generating the most student responses: playground question, brainstorm question and focal question. All three question types represented the three upper-levels of Bloom’s (1956) cognitive abilities. No subsequent research has been conducted on their use in either face-to-face or online courses.

**Purpose of study**

This study examined whether different question types influenced the quantity and quality of students’ online submissions in an undergraduate Child Development course. All discussion questions were derived from ‘hot-topic’ articles that students read in addition to the textbook and were created using the structure of one of the following six question types, based on Andrews (1980):

- **Direct link.** This version of Andrews’ (1980) playground question referred to a specific aspect of the article, such as a quotation, and asked students for their interpretation or analysis. The direct link to the article could be in the stem of the question, as in the following example, or as a requirement for the students’ response. Example: ‘“Parents are past, peers are future” (p. 124)—[specific aspect of article] Why is this quotation important to the premise about the influence of peers in one’s life? [analysis]’ Direct link questions were considered higher-order and analytical.

- **Course link.** This other version of Andrews’ (1980) playground question required specific information from the course to be integrated with a topic from the article. Example: ‘Using Family Systems Theory [course concept], explain how children’s personalities affect parenting styles [topic from article]’. Course link questions were considered higher-order and synthetic or analytical.

- **Brainstorm.** Taken from Andrews (1980), this question was structured to generate any and all ideas or solutions to an issue. Example: ‘How would you encourage [generation part] parents and teachers to emphasise praising children’s effort more than their intelligence? [issue]’ Brainstorm questions were considered higher-order and analytical or synthesis.

- **Limited focal.** This version of Andrews’ (1980) focal question presented an issue with two to four alternatives and asked students to take a position and justify it. Example: ‘Which should schools emphasise more in its curriculum [issue]: social skill development or academic skill development? [alternatives]’. Limited focal questions were considered higher-order and analytical or evaluative.
• Open focal. This version of Andrews’ (1980) focal question presented an issue with no alternatives and asked for student opinion. Example: ‘Should schools be held accountable for student performance even if it means losing funding? [issue, no specific alternatives]’. Open focal questions were considered higher-order and analytical or evaluative.

• Application. This question type was not based on Andrews (1980). It provided a scenario and asked students to respond to the scenario using information from the article. Example: ‘The Baileys have two children. One child displays a fearful personality while the other displays a stubborn personality [scenario]. What parenting style is most appropriate for each child? Justify your response [apply information from article]’. Application questions were considered lower-order and applied.

This study’s main dependent variable is higher-order thinking in student answers and responses, using Gilbert and Dabbagh’s (2005) coding scheme which defined higher-order thinking based on the upper three levels of Bloom’s Taxonomy: analysis, synthesis, and evaluation. This definition was also used in Andrews’ (1980) study and others (eg, Schrire, 2006).

Two additional dependent variables also were included for analysis: word count and degree of answer completion. Word count was included because it is a traditional manifest variable (Schrire, 2006), is easy to assess and is an objective measurement of student responses. In addition, word count may be used as a quantitative measure of ‘student participation’ (Andrews, 1980). Degree of answer completion was included as a dependent variable because many instructors traditionally grade student-written responses—whether on exams, papers, or discussions—based on whether the question was answered fully or partially. Thus, understanding which question types may lead to complete or incomplete responses can help instructors structure their online discussions. For example, if a question type that typically results in incomplete answers is chosen, the instructor will need to allot more time to help students eventually submit a complete response (or avoid the question type altogether). Finally, because word count alone does not adequately measure depth or accuracy of student learning (Schrire, 2006), this qualitative measurement was needed to see if students were addressing the issue stated in the question. Both word count and degree of answer completion were viewed as theoretically distinct from higher-order thinking and represented other possible learning outcomes for the use of online discussion.

Thus, the main research questions were:

1. Which question type generated the highest word count?
2. Which question type generated the most complete answers?
3. Which question type resulted in higher-order thinking in answers and in responses?

Method

Participants

The participants were 114 undergraduates taking a 200-level course on Child Development, taught by the primary author. Three classroom sections across two different
semesters were combined for this study. The first sample of 40 students was from the fall 2003 semester in which students attended class once a week for 2.5 hours (face-to-face) and engaged in class discussions in an online format only. The second and third samples of 74 students were from two different sections from the spring 2004 semester (37 students per section) in which students attended class twice a week for 75 minutes (face-to-face). These students participated in both in-class and online discussions. The second sample of students engaged in online discussions for the first half of the semester, then switched to in-class discussions for the second half of the semester while the third sample did the reverse.

Only students’ online responses were used for analyses; therefore, the two sections from spring 2004 were collapsed into one full semester of online discussions and added to the fall 2003 section of students (eg, responses from Section 2 students represented the first half semester of responses while Section 3 students represented the second half semester of responses). Thus, at any given time, 77 students could have responded to a given discussion (40 from Section 1 fall 2003 students and 37 from either Section 2 or 3 of spring 2004). Preliminary analyses revealed no difference amongst sections on the coded variables of word count, degree of answer completion or higher-order thinking.

Across all three sections, students ranged in age from 18–47 years ($M_{age}=20.76$ years), were predominantly female (84 females, 25 males and 5 missing data) and were predominantly Caucasian (84.2% Caucasian, 8.8% African American, 0.9% Hispanic, 0.9% Other and 5.3% missing data). Most students were in their 2nd (47.4%) or 3rd (24.6%) year of college representing 10 majors and non-degree/undecided students (top two majors: 25% Psychology and 25% Education).

Procedure
Students were required to engage in class discussion as part of their course grade and gave consent for their performance to be included in the research database. Throughout the semester, there were eight discussions of brief articles from educational news magazines on current ‘hot-topics’ in child development. Four discussions were held during the first half of the semester and four were held during the second half, at roughly equal intervals. All students were divided into eight groups of either five or six students and were kept in those groups for the full semester. Online discussions utilised the blackboard learning management system.

For each discussion, students were given three questions for a total of 24 questions in a semester. Each question represented one of the six discussion question types with any given discussion including at least two different discussion types. For example, the first discussion contained one direct link question, one open focal question and one brainstorm question. Attempts were made to have question types equally distributed across the semester.

Students were instructed that an adequate discussion score (ie, receiving a ‘B’ on the assignment) meant answering at least two of the three questions and responding to two
other group members per topic. No formal grading rubric was given to students. Online asynchronous discussions were held across one full week, with students having access 24/7. These discussions were moderated by student-research assistants who kept discussions going by contacting non-participating students and contributing reflective statements for missing group members.

Data coding and preliminary analyses
Students’ online submissions were divided into two aspects: answers to the discussion question (‘answer’) or responses to group members’ answers (‘response’). There was only one answer per question but the number of responses varied per student, from zero to four per question. Each submission was used as a unit of coding and varied in length from one sentence to several paragraphs.

The 1380 total answers and 849 total responses (2229 total submissions) were coded in three ways (dependent variables): word count, degree of answer completion and level of higher-order thinking. Word count was calculated via a computer and included each student’s answer and responses. Degree of answer completion was rated by a student assistant and the primary author on student answers only, using a 4-point Likert scale, with the lowest score representing ‘no response’ and the highest score representing ‘full response; student completely answered question’. Interrater reliability using the intraclass coefficient for the total sample was 0.96.

Coding of higher-order thinking in answers and in responses (coded separately) was rated by another research assistant and the primary author based on an adaptation of Gilbert and Dabbagh’s (2005) coding scheme (see Table 1). The categories were numbered and viewed as existing on a continuum related to Bloom’s Taxonomy. The lowest score of zero was created to code incorrect answers. The next five ratings were considered ‘lower-order thinking’ and included students: (1) only citing the readings, (2) providing their own clarification of the content, (3) bringing in prior knowledge from other resources to justify view, (4) using a real world example to explain view and (5) using an abstract example to explain view. For the last rating, which was considered higher-order thinking, students made inferences by analysing, synthesizing, or evaluating broader contextual issues when presenting their view. Seeing that the number of student responses varied, the average rating for all responses was entered into the database (eg, a student with two responses rated a one and three received the average rating of two). Each submission received only one code, the highest code earned. Interrater reliability using the intraclass coefficient for the total sample was 0.94.

This within-subjects research design required preliminary calculations to create the repeated measures variables. As each question type was represented by more than one discussion question, the average participant score per dependent variable was calculated. For example, the open focal question type was represented by four questions across four separate discussion topics (Discussion 1 Question 2, Discussion 2 Question 3, Discussion 7 Question 3 and Discussion 8 Question 1). The average word count for
those four questions was calculated to obtain the average open focal word count. This step was repeated for each question type and the series of calculations were repeated for the other dependent variables (answer completion, higher-order thinking in answers and responses).

Results

Word count
To examine which question type generated the highest word count, a repeated measures ANOVA was computed with the average word count for all six discussion type questions as the repeated measures. Results were significant, $F(5,72) = 16.004, p = 0.000, \eta^2 = 0.526$. Examination of means and follow-up targeted $t$-tests revealed three levels of rankings (for all analyses, question types within a ranking were not significantly different from one another; types between rankings were significantly different). See Table 2 for ranking based on means.

Degree of answer completion
To examine which question type generated the most complete answer, a repeated measures ANOVA was computed with the average answer completion for all six discussion type questions as the repeated measures. Results were significant, $F(5,72) = 14.005, p = 0.000, \eta^2 = 0.493$. Examination of means and follow-up targeted $t$-tests revealed four levels of rankings (see Table 2).

Table 1: Coding scheme for higher-order thinking

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Bloom’s taxonomy level</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No score</td>
<td>N/A</td>
<td>Student attempted submission but cannot be coded as a result of being too off-topic or incorrect.</td>
</tr>
<tr>
<td>1</td>
<td>Reading citation</td>
<td>1</td>
<td>Student only cited the reading using mostly direct quotations when justifying their answer.</td>
</tr>
<tr>
<td>2</td>
<td>Content clarification</td>
<td>2</td>
<td>Student stated personal interpretation of article content, such as paraphrasing ideas in own words.</td>
</tr>
<tr>
<td>3</td>
<td>Prior knowledge</td>
<td>2</td>
<td>Student used prior knowledge from class or outside resources when justifying their answer.</td>
</tr>
<tr>
<td>4</td>
<td>Real world example</td>
<td>3</td>
<td>Student applied a personal experience or scenario to justify answer.</td>
</tr>
<tr>
<td>5</td>
<td>Abstract example</td>
<td>3</td>
<td>Student applied an analogy, metaphor or philosophical interpretation to justify answer.</td>
</tr>
<tr>
<td>6</td>
<td>Making inferences</td>
<td>4, 5, 6</td>
<td>Student’s answer reflected analysis, synthesis or evaluation, made broader connections to society or culture and created new ideas in justifying answer.</td>
</tr>
</tbody>
</table>
Higher-order thinking of answers
To examine which question type generated the highest level of higher-order thinking in student answers, a repeated measures ANOVA was computed with the average thinking in answers for all six discussion type questions as the repeated measures. Results were significant, $F(5,72) = 8.279, p = 0.000, \eta^2 = 0.365$. Examination of means and follow-up targeted $t$-tests revealed three levels of rankings (see Table 2).

In order to further examine students’ higher-order thinking in answers, subsequent descriptive statistics and frequencies were computed. For all question types, the median and mode was 2. As students did not have to answer all questions for all discussions, some question types generated a greater percentage of completion than others. Table 3 lists these results.

The percentage of students receiving each of the coding levels for higher-order thinking was calculated for all six question types. Results are given below in Figure 1.

Higher-order thinking of responses
To examine which question type generated the highest level of higher-order thinking of responses, a repeated measures ANOVA was computed with the average thinking in responses for all six discussion type questions as the repeated measures. Results were significant, $F(5,72) = 6.121, p = 0.000, \eta^2 = 0.298$. Examination of means and follow-up targeted $t$-tests revealed three levels of rankings (see Table 2).

© 2007 The Authors. Journal compilation © 2007 British Educational Communications and Technology Agency.
Discussion

The purpose of this study was to discover how the structure of a discussion question (question type) influenced students’ online submissions in terms of word count, degree of answer completion and level of higher-order thinking. The six discussion question types differentially influenced students’ answers and responses on all dependent variables. The question type that was most influential depended on the dependent variable.

Question 1—Which question type generated the highest word count?

Limited focal and direct link question types generated the most words, followed by brainstorm and open focal question types. Application and course link question types generated the fewest words. These findings are similar to Andrews (1980) whose top three question types for most responses in a face-to-face discussion were playground, brainstorm and focal. In this study, Andrews’ focal question type was divided into two versions: limited and open. Focal questions asked for student opinion on some contro-
versial topic, as in a debate. Limited focal questions provided students with at least two alternatives (as in Andrews) while open focal did not specifically identify alternatives. It seemed that students had an easy time generating words when asked for their opinion, especially if they were given alternatives.

Andrews’ (1980) playground question was also divided into two versions: direct link and course link, with the direct link type most similar to his original definition. Course link questions required students to bring in material from the class and it appeared that students had a harder time with that task than merely providing an interpretation of one aspect of the article, as required in direct link questions. Brainstorm questions were helpful in soliciting quantity in both the current study and Andrews’ because this type requested any and all ideas from students, it may have freed them to write a significant amount of material. Finally, the newly created application question, along with the aforementioned course link type, generated the lowest word count. Perhaps students viewed these two question types differently from the others. Students refrained from answering the application and course link question types most (see Table 3), possibly indicating their dislike for the structure of those question types. Future research is needed to replicate these findings and to assess why students chose these question types the least compared to the other types.

**Question 2—Which question type generated the most complete answers?**

Limited and open focal question types generated the most complete answers, followed by brainstorm type and then direct link type. Application and course link question types generated the least complete answers. These findings were similar to the results from the first research question, with only the direct link and open focal question type switching rankings. Logically, word count and degree of answer completion would be highly correlated. Incomplete answers tended to have fewer words and complete answers tended to have many words, although there were exceptions (e.g., a correct and concise answer).

Students were likely able to give more complete answers to the two focal question types because of the opinion-oriented nature of these questions. As long as students thoroughly explained their views, they received full credit. This reasoning also may explain how students were able to provide their ideas more completely for the Brainstorm questions than the remaining three types. Direct link, applications and course link required students to respond in a particular manner. If students neglected to address the specific part of the article, could not respond to the scenario or did not adequately bring in material from the course, their degree of answer completion rating was lower.

**Question 3—Which question type resulted in higher-order thinking in answers and responses?**

For answers, course link, brainstorm and direct link question types resulted in higher levels of thinking than limited focal type. Open focal and application question types resulted in the lowest level of thinking, based on Bloom’s (1956) Taxonomy. For
responses, course link, brainstorm and open focal resulted in higher levels of thinking than limited focal and direct link types. Application type resulted in the lowest level of thinking.

Whereas course link questions were least influential for word count and answer completion, they were most influential for higher-order thinking. This question type required students to bring in prior knowledge or outside resources, which if done successfully, resulted in a rating of a least a three on the coding scheme (see Table 1). Indeed, this question type generated the highest percentage of three’s (see Figure 1). Thus, structuring a question to require synthesis of material can assist a student in thinking in more complex ways. Future research could explore other question types that require students to answer using a real world or abstract example to see if their thinking level improves. For example, researchers may want to include the question stems created by King’s (1992) ‘guided cooperative questioning’ research. These question stems have been used successfully for in-class discussions related to lecture comprehension or problem-solving and may transfer to the context of an online discussion.

Brainstorm question type also ranked highly for both answers and responses. This question type seemed to facilitate students justifying their solutions by bringing in prior knowledge or examples from their own life (see Figure 1). Similar findings occurred for direct link question type for students’ answers but not for responses. These top three question types (course link, brainstorm and direct link) were the only ones to solicit any inferences (code rating of six, higher-order thinking) from students. When students answered questions associated with the two focal types or application, they tended not to justify them with outside resources, examples or inferences. These students may have thought their opinion alone was good enough without providing additional support in their reasoning.

This study was one of the first to examine undergraduates’ level of thinking using a coding scheme based on Bloom’s Taxonomy. Overall, their level of thinking was lower-order, with all means, medians and modes associated with Bloom’s (1956) comprehension level. This finding is inconsistent with Schrire’s (2006) study in which many graduate students engaged in higher-order thinking. This inconsistency could be a result of the differences in methodology or, more likely, participant samples, as undergraduates in a 200-level course are expected to be different from graduate students. One suggestion for increasing undergraduate students’ level of thinking is to provide them with more structure for the discussion, as Gilbert and Dabbagh (2005) found when facilitator guidelines and evaluation rubrics were presented to graduate students. Future research could explore the possibility of providing undergraduates with a scoring rubric and/or training on how to make inferences in online responses in order to see improvements in their frequency of engaging in higher-order thinking.

In addition, future studies may want to require students to answer all discussion questions, instead of giving them the option of answering only some. Students from this study needed to answer just two of the three questions and respond to two other group
members per topic to receive a ‘B’ on the assignment. If students are required to respond to every discussion question, it is possible that the ranking of these question types may be different for all dependent variables, particularly word count.

Conclusion
Instructors wanting to improve the quantity and quality of online discussions could structure their discussion questions using the successful question types found in this study. If the learning goal is to facilitate a large quantity of discussion whereby students are giving fairly complete answers to questions, then using the limited focal type would be the best strategy, followed by brainstorm, open focal and direct link. If the learning goal is to facilitate students’ level of higher-order thinking, then the course link, brainstorm and direct link questions would work best. The application question type scored lowest for all dependent variables; however, additional research is needed to understand why this finding occurred.

This study was statistically strong in that it used a within-subjects design, analysed a large number of online submissions and found results that were highly significant with moderate effect sizes. There were, however, some limitations. The sample of students was predominantly Caucasian and female; thus, generalisations to more diverse samples may be limited. Each question type was represented by at least three discussion questions but some types had more questions associated with them than others. Future research should make sure that there are equal numbers of questions per discussion type. Despite these limitations, this study provides instructors with guidance on how to create successful questions to improve online discussion.

References


