Imagine you are working on a research paper about the **increase of technology in education and online learning**. Read the three information sources that follow this page and keep the CAARP model in mind as you review each source.

*Remember:*  
C = Currency  
A = Authority  
A = Accuracy  
R = Relevance  
P = Purpose

For the third and final source you will see the address (URL) of a website. Click on that link to be taken to a website. Please review the website as a whole for your third and final source.

To complete your assignment, go to: [http://library.uncw.edu/instruction/UNI_library_assignment](http://library.uncw.edu/instruction/UNI_library_assignment). Login at the bottom of the page and follow the directions to answer questions about each information source.
Many schools across the nation are working to include online learning as a fundamental part of the collection of teaching strategies. In 2000, approximately 98% of the public schools in the United States were connected to the Internet (National Center for Education Statistics, 2001) and were working to expand their students’ communication options from the classroom to a variety of online discourse opportunities with peers, community members, and content experts (Palloff & Pratt, 1999; Wegerif, 1998). This action-research pilot study was designed to answer three questions: Which of two methods, the Newman or the Facione, most reliably assesses students’ critical thinking during in-class and online discussion? How do the methods compare in consistency of use? How do the methods compare in ease of use?

One of the most powerful tools for engaging students in dialogues with each other and with the larger society is the online discussion board, a form of computer-mediated communication (CMC). With the burgeoning adoption of such course management tools as Blackboard, FirstClass, and WebCT (Blackboard, 2004; FirstClass, 2004; Goldberg, 2000), more schools have access to sophisticated online discussion tools and are asking questions about how this type of discourse affects student learning and thinking.

Over the past decade, some researchers have speculated that there may be a relationship between the use of online discussions and the development of critical thinking skills (Garrison, 1992; MacKnight, 2000; Scarce, 1997). One hypothesis is that CMC provides a social context for learning that gives learners time to think about their contributions and organize their thoughts prior to responding.

Most of the studies examining the value of online discussion or CMC admit that more work is needed to understand the complexities of online discussions, to document how they differ from in-class discussions, and to develop valid and reliable tools for qualifying behavior (Love, 2002). Research into the complexities of online discussion has explored the role of the instructor and the wording of the discussion task (Tu & McIsaac, 2002). Ou, Ledoux, and Crooks (2004) reported that the presence of the instructor during CMC had a positive impact on the students’ evaluation and analysis of ideas but not on their connection of ideas. MacDonald and Caverly (2001) discussed three types of online discussion and a five-step model for leading students through various types of CMC discussion. Greenlaw and DeLoach (2003) described how the wording of a discussion topic can affect levels of critical thinking. They also observed that CMC discussions started more slowly than face-to-face discussions and were more disjointed in organization and flow. Rodrigues (1999) found that her graduate students’ online discussions were more detailed and reflective.

The most frequently used tools for qualifying student behavior during class activities involve some form
of content analysis. These procedures can take the form of discrete skill analysis (Newman, Johnson, Cochrane, & Webb, 1996; Newman, Webb, & Cochrane, 1995) or holistic analysis (Facione & Facione, 1994; Greenlaw & DeLoach, 2003) of student discourse. The preponderance of research studies have occurred at the postsecondary level.

**PURPOSE**

This study evolved as an action research project. Secondary school administrators who shared an interest in assessing their use of online learning tools contacted the researchers. As a result, a collaborative study was designed to answer three questions:

1. Which of two methods, the Newman or the Facione, most reliably assesses students’ critical thinking during in-class and online discussions?
2. How do the methods compare in consistency of use?
3. How do the methods compare in ease of use?

**METHOD**

**Participants**

The group of students whose discussions were analyzed attended a midwestern metropolitan high school. They were enrolled in an American Government AP Honors course. The school was known for its rich infusion of modern technologies, so the teacher and the students were accustomed to using computers frequently. The school provided Blackboard course management software for all classes. The American Government teacher involved in this study had frequently used the Blackboard Discussion Board.

**Analytical Tools**

In the search for valid and reliable tools for evaluating levels of critical thinking, the researchers located two methods that appeared to have face and construct validity. Both are based on theories that are commonly accepted in research literature. The research project was designed to compare in-class and online discussions using these two methods. A description of those methods follows.

**The Newman Method**

The Content Analysis Method to Measure Critical Thinking and Computer Supported Group Learning was developed by Newman, Webb, and Cochrane of Queen's University Belfast (1995). A series of papers report the researchers’ analysis of face-to-face and asynchronous, online discussions (Newman, 1994; Newman, Johnson, Cochrane, & Webb, 1996; Newman, Webb, & Cochrane, 1995; Webb, Newman, & Cochrane, 1994). Their conceptual framework was based on Garrison’s (1992) description of critical thinking as the construction of meaning through internal reflection by the individual and the sharing of personal constructs, thereby establishing a “cognitive presence” in the discourse. Garrison further described the collaborative validation of meaning through the sharing of ideas in group discussion, thus forming a “community of learning” that promotes higher level thinking (Garrison, 1992).

Garrison (1992) perceived the abilities used in critical thinking as similar to the skills involved in the problem-solving process:

1. Problem identification skill
2. Problem definition skill
3. Problem exploration skill
4. Problem evaluation skill
5. Problem integration skill

Using Garrison’s (1992) five-stage model of critical thinking, the Newman team developed a coding technique to categorize the content of student iterations using a dichotomous pairs approach suggested by Henri (1991). The dichotomous pairs became their criteria to accurately assess the range of critical thinking occurring during discussions. These criteria, along with the negative and positive descriptors, are shown in Table 1. Using the Newman scale, raters assigned a positive or negative code each time they recognized a behavior described by the Newman criteria. For some student remarks, several codes could be assigned. Once the codes were assigned, raters tallied the number of positive codes and the number of negative codes. The following equation was applied to arrive at a “critical thinking (CT) ratio” for each indicator (Newman, Johnson, Cochrane, & Webb, 1996):

\[(\text{EQ}) \quad \text{Depth of CT ratio} = \frac{(x+ - x-)}{(x+ + x-)}\]

Thus, the result is a ratio for each indicator that ranges from a $-1$ to a $+1$, representing the relationship of positive to negative ratings.
The Facione Rubric

A diametrically different approach is used in the Facione model for assessing critical thinking behaviors. Drs. Peter and Noreen Facione developed and copyrighted a “Holistic Critical Thinking Scoring Rubric” (Facione & Facione, 1994; available online at http://www.insightassessment.com/HCTSR.htm). The rubric, rating form, and instructions are available for unlimited duplication for “local teaching, assessment, research or other educational and noncommercial uses,” provided the authors are cited. The Facione rubric uses much broader criteria than the Newman Method to describe critical thinking. The authors contend that the following six behaviors are involved

### TABLE 1. Newman Team Indicators of Crucial (+) and Uncritical (–) Thinking

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance</td>
<td>R+ relevant statements</td>
<td>R– irrelevant statements, diversions</td>
</tr>
<tr>
<td>Importance</td>
<td>I+ important points/issues</td>
<td>I– unimportant, trivial points/issues</td>
</tr>
<tr>
<td>Novelty: new information, ideas, solutions</td>
<td>NP+ new problem-related information</td>
<td>NP– repeating what has been said</td>
</tr>
<tr>
<td></td>
<td>NI+ new ideas for discussion</td>
<td>NI– false or trivial leads</td>
</tr>
<tr>
<td></td>
<td>NS+ new solutions to problems</td>
<td>NS– accepting first offered solution</td>
</tr>
<tr>
<td></td>
<td>NQ+ welcoming new ideas</td>
<td>NQ– squashing, putting down new ideas</td>
</tr>
<tr>
<td></td>
<td>NL+ learner (students) brings new things in</td>
<td>NL– dragged in by tutor</td>
</tr>
<tr>
<td>Bringing outside knowledge or experience</td>
<td>OE+ drawing on personal experience</td>
<td>QQ– squashing attempts to bring in outside knowledge</td>
</tr>
<tr>
<td></td>
<td>OC+ referring to course material</td>
<td>O– sticking to prejudice or assumptions</td>
</tr>
<tr>
<td></td>
<td>OM+ using relevant outside material</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OK+ evidence of using previous knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OP+ course-related problems brought in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NL+ learner (students) brings new things in</td>
<td></td>
</tr>
<tr>
<td>Ambiguities: clarified or confused</td>
<td>AC+ clear, unambiguous statements</td>
<td>AC– confused statements</td>
</tr>
<tr>
<td></td>
<td>A+ discussing ambiguities to clear them up</td>
<td>A– continuing to ignore ambiguities</td>
</tr>
<tr>
<td>Linking ideas, interpretation</td>
<td>L+ linking facts, ideas, and notions</td>
<td>L– repeating information without making inferences or offering an interpretation.</td>
</tr>
<tr>
<td></td>
<td>L+ generating new data from information collected</td>
<td>L– stating that one shares the ideas or opinions stated without taking these further or adding any personal comments.</td>
</tr>
<tr>
<td>Justification</td>
<td>JP+ providing proof or examples</td>
<td>JP– irrelevant or obscuring questions or examples</td>
</tr>
<tr>
<td></td>
<td>JS+ justifying solutions or judgments</td>
<td>JS– offering judgments or solutions without explanations or justification</td>
</tr>
<tr>
<td></td>
<td>JS+ setting out advantages and disadvantages of situation or solution</td>
<td>JS– offering several solutions without suggesting which is the most appropriate.</td>
</tr>
<tr>
<td>Critical assessment</td>
<td>C+ critical assessment/evaluation of own or others’ contributions</td>
<td>C– Uncritical acceptance or unreasoned rejection</td>
</tr>
<tr>
<td></td>
<td>CT+ tutor prompts for critical evaluation</td>
<td>CT– tutor uncritically accepts</td>
</tr>
<tr>
<td>Practical utility (grounding)</td>
<td>P+ relating possible solutions to familiar situations</td>
<td>P– discussing in a vacuum (treat as if on Mars)</td>
</tr>
<tr>
<td></td>
<td>P+ discussing practical utility of new ideas</td>
<td>P– suggesting impractical solutions</td>
</tr>
<tr>
<td>Width of understanding (complete picture)</td>
<td>W+ widening discussion (problem within a larger perspective; intervention strategies within a wider framework.)</td>
<td>W– narrowing discussion (address bits or fragments of situation; suggest glib, partial interventions)</td>
</tr>
</tbody>
</table>

when a good critical thinker arrives at “a purposive judgment”:

1. analysis,
2. interpretation,
3. evaluation,
4. inference,
5. explanation, and
6. metacognitive self-regulation.

Using these descriptors, the Faciones constructed a fourth-level performance-based rubric for analyzing critical thinking behaviors, with 4 being the highest and 1 the lowest level of performance (see Table 2). The Faciones also spell out requirements for the raters using the instrument. Raters must “differentiate and focus” to achieve success and must apply “critical thinking, content knowledge, and technical skill (craftsmanship)” to the rating process. The authors recommend a period of training and practice for raters to achieve “consistency and fairness.” Through practice, coordination, and reconciliation, the Faciones have found that raters can achieve levels of reliability necessary for meaningful application (Facione & Facione, 1994).

It is not clear whether this instrument is intended to evaluate both individual performances and group discussions. The instructions state, “Usually, two raters will evaluate each essay/assignment/project/performance.” At any rate, it is expected that there will be one score for the entire body of discourse and that the raters reconcile differences by (a) mutual conversation between the two raters, (b) using an independent third rater, or (c) taking the average of the two initial ratings. The averaging strategy, however, is strongly discouraged by the authors.

| Table 2. Holistic Critical Thinking Scoring Rubric (Facione & Facione, 1994) |
|-----------------|-----------------------------------------------------------------------------|
| **Level** | **Description** |
| 4 | **Consistently does all or almost all of the following:**  
Accurately interprets evidence, statements, graphics, questions, etc.  
Identifies the salient arguments (reasons and claims), pro and con.  
Thoughtfully analyzes and evaluates major alternative points of view.  
Draws warranted, judicious, non-fallacious conclusions.  
Justifies key results and procedures; explains assumptions and reasons.  
Fair-mindedly follows where evidence and reason lead. |
| 3 | **Does most or many of the following:**  
Accurately interprets evidence, statements, graphics, questions, etc.  
Identifies relevant arguments (reasons and claims), pro and con.  
Offers analyses and evaluations of obvious alternative points of view.  
Draws warranted, non-fallacious conclusions.  
Justifies some results or procedures; explains reasons.  
Fair-mindedly follows where evidence and reason lead. |
| 2 | **Does most or many of the following:**  
Misinterprets evidence, statements, graphics, questions, etc.  
Fails to identify strong, relevant counterarguments.  
Ignores or superficially evaluates obvious alternative points of view.  
Draws unwarranted or fallacious conclusions.  
Justifies few results or procedures; seldom explains reasons.  
Regardless of the evidence, maintains or defends views based on self-interest or preconceptions. |
| 1 | **Consistently does all or almost all of the following:**  
Offers biased interpretations of evidence, statements, graphics, questions, information, or the points of view of others.  
Fails to identify or hastily dismisses strong, relevant counter-arguments.  
Ignores or superficially evaluates obvious alternative points of view.  
Args using fallacious or irrelevant reasons, and unwarranted claims.  
Does not justify results or procedures, nor explain reasons.  
Regardless of the evidence, maintains or defends views based on self-interest or preconceptions.  
Exhibits close-mindedness or hostility to reason. |
Every effort was made by the evaluation team participating in this study to adhere to the guidelines provided by the developers of both methods (Facione & Facione, 1994).

**Procedure**

The American Government discussions collected for analysis during this study consisted of two sets of questions, each including a classroom and online discussion. The teacher provided four questions, two per topic. Considerable effort was made to generate questions that were comparable in complexity. Because the research team felt that the order of the discussions could influence the discussion, the first set of questions was done with the in-class discussion first and the online discussion last. For the second set of questions, the online discussion was first and the in-class discussion last. The in-class discussions were audiotaped and transcribed for analysis. The online discussions were archived and printed. All student names were replaced with pseudonyms.

The teacher was given the freedom to structure the discussions in the usual manner. Because the class met for 2 hours on Wednesday, the online discussion was held either the week prior to the in-class discussion or the week following in-class discussion. This meant that students had a week to discuss their questions online.

The students also had time in class to discuss the questions. The teacher divided the students into groups for the in-class discussions, but had them discuss as a whole group when online. Table 3 depicts the questions asked for each set, the order, and the groupings used for each set of questions.

After the discussions were transcribed, they were distributed to a three- or four-person team of educators for analysis using each of the methods discussed above. After individual assessment, the evaluators met as a team to clarify the methods and compare results. In some cases, second- and third-round assessments were conducted after terms or approaches were clarified or agreed on by the team.

As a follow-up to the rating sessions, the research team independently responded to a brief questionnaire soliciting their opinions about the rating methods. The questionnaire items include the following:

1. How easy was this tool to use?
2. How much did this tool expand your understanding of critical thinking?
3. How accurately did you feel it captured the level of thinking (validity)?
4. How comprehensive/thorough did you think it was?
5. Other criteria? Comments?

<table>
<thead>
<tr>
<th>Set</th>
<th>First question</th>
<th>First discussion</th>
<th>Second question</th>
<th>Second discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>You are a member of an illustrious committee that has been commissioned to nominate presidents to various categories in the U.S. Presidents’ Hall of Fame. You have been asked only to consider presidents of the 20th century. Categories include Exhibited Greatest Character and Best Policy Maker/Shaper. Things to consider: subcategories, specific examples to support or negate individual nominations. Use lots of sources. Hand in 1 page of notes on each category.</td>
<td>• In-class discussion</td>
<td>You are a member of an illustrious committee that has been commissioned to nominate presidents to various categories in the U.S. Presidents’ Hall of Fame. You have been asked only to consider presidents of the 20th century. Categories include exhibited greatest leadership and delivered on Campaign Promises. Things to consider: subcategories, specific examples to support or negate individual nominations. Use lots of sources. Hand in 1 page of notes on each category.</td>
<td>• Online discussion</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>Small Group A &amp; Small Group B</td>
<td>Online discussion</td>
<td>Whole group</td>
</tr>
</tbody>
</table>
RESULTS

Question 1
Which of two methods can best describe levels of critical thinking that occur during in-class and online discussion?

As described above, the two methods being scrutinized in this study took very different approaches to the definition of critical thinking and the identification of critical thinking behaviors during discussion. Because the Newman method involved the identification of several different attributes that could operate simultaneously, raters had trouble keeping all criteria in mind at any given time. Therefore, the resulting tallies, especially when considered one attribute at a time, were quite different among the raters. The raters’ final ratios also were inconsistent. Several attempts were made to attain interrater reliability, but although the reliability improved, consistency was not achieved.

The Facione method, by contrast, was general to a fault. There were only four levels described by the rubric. Because the discussions being evaluated were extensive, raters often felt that the level changed throughout the sample. Although there was greater agreement on what the ratings for the total samples would be, the data were insufficient to compare and contrast among samples. So, again, the team was unable to achieve the results desired.

In general, the two methods did help the raters identify critical thinking behaviors but not to the degree of reliability needed to perform a reliable research study. They were able to see that all discussions included statements that demonstrated identification of relevant arguments; analysis and interpretation of pros, cons, and alternative viewpoints; justification for opinions; and appropriate conclusions. The raters agreed that neither of these two tools, however, proved to be stable enough for continued use.

Question 2
How do the methods compare for consistency of use? To gain experience with each method, the rating team agreed to select a sample of 100 utterances from each discussion. All samples occurred toward the end of the discussions in the hopes that higher levels of analysis had been built as the interaction progressed. The rating team members hoped to gain experience evaluating as many different levels of thinking as possible.

Newman Method. In addition to the problems with the Newman method described above, the four raters could not agree on how much of a sample should constitute a unit of evaluation. Nor was there any guidance provided by the method’s authors to help with this dilemma. Three of the raters opted to rate each speech in the discussion. The fourth rater chose to rate “chunks” of meaning. Unfortunately, the raters did not achieve interrater reliability to a satisfactory level (\(r = -0.09\) to \(0.42\), \(p \geq 0.05\); see Table 4). They discussed their rationale and decided to try rating again. This time reliability was calculated for only the three raters coding at the sentence level of discussion. Again, the interrater reliability scores fell short for these samples (\(r = -0.35\) to \(0.62\), \(p \geq 0.05\); see Table 5).

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pearson correlation</td>
<td>1.000</td>
<td>0.437</td>
<td>-0.0999</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.207</td>
<td>0.785</td>
<td>-0.233</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Pearson correlation</td>
<td>0.437</td>
<td>1.000</td>
<td>0.000</td>
<td>-1.870</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.207</td>
<td>0.000</td>
<td>0.000</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Pearson correlation</td>
<td>-0.099</td>
<td>0.000</td>
<td>1.000</td>
<td>0.442</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.785</td>
<td>1.000</td>
<td>0.000</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Pearson correlation</td>
<td>0.423</td>
<td>-0.187</td>
<td>0.442</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.223</td>
<td>0.605</td>
<td>0.201</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
**Facione Method.** The researchers independently rated both sets of discussion questions using the holistic rubric. When they met to compare their ratings, it was apparent there was too much variance between the ratings. As advised in the rubric’s instructions, the group stopped and collectively answered common questions. One problem that arose for all raters was, “There are so many levels of thinking displayed during the course of these long discussions, how do we know which to assign to the whole?” The group decided to review their ratings and assign the highest level that appeared to have persisted for considerable duration during the discussion.

The group then scored both in-class and online discussions one more time using the Facione Rubric. The group achieved greater agreement on this round of ratings (see Table 6). Because there was only one score for each sample, however, there were insufficient data to run reliability tests. Due to the nature of holistic ratings, raters used a percentage of agreement rather than the statistical procedure used for the Newman method.

Raters questioned the reliability of both instruments. When asked how accurately they felt that each captured the level of critical thinking, two of the raters felt the Newman was stronger, one felt neither accomplished this goal, and one felt the Facione was slightly stronger. All raters became frustrated trying to achieve some consistency using the Newman. One felt she got less consistent with the Newman across time; another felt she repeatedly “regressed to the mean of ‘3’” on the Facione.

Thus, each of the instruments had problems with reliability. The complexity of the Newman method made it difficult to keep all criteria in mind at all times during the rating process. Because the four Facione levels were general in nature, comparison ratings and samples were not reliable.

### TABLE 5. Interrater Reliability Correlations Using the Newman Method for Online Discussions

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pearson correlation</td>
<td>1.000</td>
<td>-0.254</td>
<td>-0.354</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.544</td>
<td>0.436</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Pearson correlation</td>
<td>-0.254</td>
<td>1.000</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.544</td>
<td>0.000</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Pearson correlation</td>
<td>-0.354</td>
<td>0.620</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.436</td>
<td>0.101</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

**Question 3**

*How do the methods compare in ease of use?* As described above, the research team independently responded to a brief questionnaire soliciting their opinions about the rating methods at the close of the study. Their responses revealed more about the two instruments. When asked about the ease of use, all agreed that the Newman method, because of its multiple indicators, was more time-consuming than the Facione Rubric. The time required to use the Newman method was so demanding that the raters felt it was impractical to use over many samples. The Facione, although less time-consuming, was “too vague” due to the generality of the criteria to use with the extended discussions in this study.

**DISCUSSION**

The search for a valid and reliable tool that is easy to use for assessing levels of critical thinking in high school students’ discussions was the impetus for this study. In a collaborative effort, high school administrators and university researchers sought to compare the validity and reliability of two analytical tools for evaluating high school student critical thinking.

The Newman method appears to be too cumbersome for use with long discussions. The difficulties with this method related to training and time. McKlin, Harmon, Evans, and Jones (2002) took three rater reliability trials before reaching an acceptable level of consistency for use of the Newman method. Although Newman and colleagues reported acceptable levels of rater reliability, they did not describe their rater training process in great detail. Nor did they offer guidance on the length of each unit of analysis. This length of sample appears to be an important issue in determining the practicality of the tool. Sherry, Gavalin, and Billing (2002) questioned how much of a given sample is needed to make sense of the nature of critical thinking that is occurring. McKlin and colleagues (2002) advocated for a controlled word count to make the analysis task more manageable.

A second problem associated with the Newman method concerned the time involved to analyze student discussions. The amount of time required to train and then code using the method’s extensive descriptors far exceeded that available for the average professional in a school setting. Raters complained that it took several times through the sample to consider all criteria throughout the sample. This concern also was expressed by the Newman group, indicating that more time was needed to gain consistency among ratings.
It appears as if the Facione holistic method did lend itself better to higher levels of reliability and easier application to large samples of discussion. There were two problems associated with this method, however. One was that in longer units of study, the varying levels of critical thinking that were evident during the course of the discussion made it difficult to assign an overall rating. Another was, with only one number for a whole discussion, statistical formulas could not be used for comparisons between ratings and samples.

The raters disagreed on how comprehensive or thorough they judged each instrument to be. All indicated they had learned from both instruments, however. When the group was asked how much the tool expanded their understanding of critical thinking, most felt that each of the instruments accomplished this in its own way.

**Recommendations**

In light of the increasing reliance on online discussion in education settings, more research is needed to guide educators who use this medium for teaching and learning. Tools to measure outcomes are essential if this form of discourse is to be established as a viable medium for developing critical thinking. Sample size (total discussion vs. portions only), size of core sample for analysis, target audience (student or group), presence of an instructor, and time feasibility need further investigation. Some work has been done at the Georgia State University on the use of neural networks to analyze critical thinking during online discussions (McKlin et al., 2002). The use of these tools would save hours of work spent training raters and coding discussion samples, but this approach also depends on more reliability in assessment instruments.

Most educators who have worked extensively with online education realize that determining whether this form of learning can have a positive effect on students’ levels of critical thinking is an important area of inquiry. Given the attention that higher level thinking has been given in educational theory and practice throughout the years, it is surprising that more reliable tools are not available for assessing this highly desired outcome of learning.

It is the recommendation of the researchers involved in this study that more work be done and more sharing take place to refine tools to evaluate this commonly desired learning outcome—critical thinking.

**ABOUT THE AUTHORS**

Melodee Landis, PhD, is an associate professor of teacher education in the College of Education at the University of Nebraska at Omaha. Her interests include critical thinking, instructional technology, and the lack of diversity in the teaching field. She teaches human development and learning, and instructional design for integrating technology. Kristine D. Swain, PhD, is an assistant professor of special education at the University of Nebraska at Omaha. Her current research interests include language literacy, ongoing assessments, and motivational strategies for students with disabilities. Mary J. (Morris) Friehe, PhD, is a full professor in the Department of Special Education and Communication Disorders, College of Education, at the University of Nebraska at Omaha. Her expertise is in the areas of language sample analysis and language-based learning disabilities. She teaches classes on school-age language disorders and multicultural issues for speech–language pathologists. Kathy L. Coufal, PhD, is chair of the Department of Communication Sciences and Disorders at Wichita State University. Her current research interests include early language development and disorders, emergent literacy, and family-centered intervention. She teaches doctoral seminars in psycholin-
guistic theory and single-subject research design. Address: Melodee Landis, Kayser Hall 522, University of Nebraska at Omaha, Omaha, NE 68182-0163.

REFERENCES
Ou, C., Ledoux, T., & Crooks, S. M. (2004). The effects of instructor presence on critical thinking in asynchronous online discussion. In C. Crawford et al. (Eds.), Proceedings of Society for Information Technology and Teacher Education International Conference (pp. 2989–2993). Chesapeake, VA: AACE.
Technology will change education

Author: Clemens, Jason; Frazier Fathers

Abstract:
The use of this technology is likely to expand in the U.S. as the Bill and Melinda Gates Foundation has announced tens of millions of dollars in grants to support similar technologies across the United States.

Full text:

There is almost no aspect of life that has not been fundamentally changed and reordered over the past 100 years through new technology and innovation. Think for instance, of how we communicate, travel, work, and even rest, and you’re hard-pressed to find one aspect of life that has not been materially altered by technology.

Yet there is one significant aspect of society where very little has changed over the last century: education. It is the one area where a worker, in this case a teacher, from 1913 could be transported to 2013 and adapt quite easily to the modern world.

For all intents and purposes, we educate our children in much the same way we did a century ago, in a one-size-fits-all manner. Specifically, most of our education system still relies on a teacher (or professor) formally instructing students in a classroom setting.

Technology is poised to change this by revolutionizing the learning process. The development and quickly emerging technology of interactive adaptive software holds the potential to dramatically change how we think about and deliver education. Adaptive learning software tailors the learning experience to individual students in terms of both their strengths and weaknesses.

This enables students to individually progress through material at a pace commensurate with their understanding and aptitude in a particular subject matter rather than the current situation where some students are left behind while others are bored because the lessons are delivered in a one-size-fits-all format. The key to this individual education is the adaptive nature of the technology. As students work through problems, material, and even gamelike simulations, the software monitors and assesses their answers and response times.

In addition, such systems constantly provide teachers with real-time results for each student. This allows teachers to more frequently monitor how students are progressing and intervene when needed. It allows teachers to become more effective individual coaches for students.

Using such technology, teachers can now engage their students in a more personalized, individual manner. It is not simply a theoretical exercise. There are a host of real-world examples of innovative educators and schools employing such technology to improve education for students.

One of the best known is the Khan Academy in the United States. Students using this system review lessons and material through online videos outside of the classroom and then engage with teachers on the materials in the classroom. This approach is meant to focus in-class time on areas where students need assistance.

The use of this technology is likely to expand in the U.S. as the Bill and Melinda Gates Foundation has announced tens of millions of dollars in grants to support similar technologies across the United States.
Recently, McGraw-Hill Education brought this cutting-edge technology to Canada with SmartBook. These first of their kind adaptive e-books are able to modify material and questions based on student responses.

There are hurdles to this new technology, however, not the least of which is a better understanding of its potential and how best to employ it in the education system.

There are also practical barriers. Many teachers remain opposed to such technology simply as a function of the degree to which it can change the status quo. School boards and provincial governments also need to better understand the possibility of such technology and begin implementing policies that allow for experimentation and reform rather than protecting the status quo.

We very well may be at the cusp of a revolution in education based on technological change. Preliminary indications are that emerging technologies can markedly improve many of the problems observed in education. While challenges and barriers remain, understanding the potential of this new technology is paramount to secure policies that facilitate experimentation and innovation in education. The possibility of vastly improving education is real and immediate.

Jason Clemens and Frazier Fathers are co-authors of Education and Technology: A Primer, available at fraserinstitute.org. This column was distributed by Troy Media. troymedia.com
Click on the link below. Examine the website and answer the questions for “Source 3.”

https://www.tie.net/