Harnessing Collaborative Annotations on Online Formative Assessments

Jian-Wei Lin1* and Yuan-Cheng Lai2

1Department of International Business, Ching Yun University, Taoyuan, Taiwan // 2Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan // jwlin@cyu.edu.tw // laiyc@cs.ntust.edu.tw

*Corresponding author

(Submitted July 07, 2011; Revised October 05, 2011; Accepted October 31, 2011)

ABSTRACT

This paper harnesses collaborative annotations by students as learning feedback on online formative assessments to improve the learning achievements of students. Through the developed Web platform, students can conduct formative assessments, collaboratively annotate, and review historical records in a convenient way, while teachers can generate formative assessments, manage collaborative annotations, and spot blockades to student learning in an efficient way. The experiment spanned one semester and underwent midterm and final exams. The results showed that adopting collaborative annotations on formative assessments significantly enhanced the learning achievements of students and extended retention time in the proposed system. Through a questionnaire and interviews, most students expressed positive attitudes toward using the proposed system.

Keywords

Online formative assessment, Collaborative annotation

Introduction

Formative assessments, continuously embedded in the teaching and learning process of a curriculum, attempt to improve learning achievements by offering feedback in the process. Summative assessments are used to check learning achievements at the end of the curriculum (Bell & Cowei, 2001). The point when the assessment occurs in a curriculum distinguishes these two types of assessments. The objective of formative assessments is to provide feedback to students rather than to evaluate them for course grades. Such feedback used in educational contexts is generally regarded as crucial for improving learner knowledge and skill acquisition (Pridemore & Klein, 1991). The feedback also gives learners the opportunity to develop their cognitive strategies and to rectify misconceptions through the training period (Azevedo & Bernard, 1995).

After teaching the course “Electrical Commerce” numerous times, it became clear to the author that students prepared for midterm or final exams at the last minute instead of ongoing preparation, leading to the poor preparation of students to take exams and the inadequate ability to evaluate their own knowledge about course material. Moderate adoption of formative assessments in the course drives students to practice course material in their daily lives. However, even given an adequate online formative assessment system, low participation rate among students significantly influences learning effects and in turn downgrades learning achievements. Costal (2010) indicated low participation rate as a critical concern to an online formative assessment in tertiary education. Henly (2003) reported that low ranking students, who are most likely to benefit from using an e-Learning system, were less likely to fully use online resources. Similarly, in the research of Buchanan (2000), only one tenth of undergraduate students had ever used an online formative assessment tool during a semester. Motivating students to use online formative assessment tools is an essential issue and designing appropriate methods to motivate students to complete formative assessment tasks is an important direction for future research in this area (Costal et al., 2010).

Web technology has proven that online collaborative annotations can improve both learning achievements and motivation. However, previous works have mainly focused on collaboratively annotating electronic materials or documents (Su et al., 2010; Yang et al., 2004; Yeh & Lo, 2009). Works discussing the learning performance and behavior of students when harnessing collaborative annotations in the process of online formative assessments is scarce.

Sustaining continuous formative assessment activities in a curriculum in a traditional learning environment is also time-consuming for a teacher. This is because teachers must prepare test content, evaluate student performance, identify student- learning barriers, and provide appropriate feedback. Teachers often encounter many students...
simultaneously and face pressure to adhere to scheduled instructional progress; therefore, it becomes more difficult for teachers to effectively administer formative assessments and offer appropriate feedback (Bransford et al., 2000; Wang, 2008). An efficient online formative assessment could help teachers promote teaching quality and student learning efficiency.

This study first uses collaborative multimedia annotations as feedback for online formative assessments, to raise student participation rate and to enhance student summative assessments. Collaborative annotations are used to scaffold learning to provide the necessary information to correctly answer particular questions. The proposed system can then facilitate both student learning and teacher management in formative assessments. Using the developed Web platform, students can conveniently conduct formative assessments, collaboratively annotate, and review historical records, and teachers can efficiently manage formative assessment activities, and instantly spot blockades to student learning to adjust teaching strategies. The system is currently available at the web address: http://ec1.it.cyu.edu.tw/AppCertForEC/Student/user_login.aspx in the Chinese version.

The experiment lasted for one semester and underwent midterm and final exams. The results showed that adopting collaborative annotations on formative assessments significantly enhanced student learning achievement and substantially extended student retention time of the proposed system. Most students expressed positive attitudes toward using the proposed system through questionnaires and interviews.

The rest of this paper is organized as follows: Section II introduces related literatures. Section III elaborates the proposed approach and system. Section IV presents the experimental results. The conclusion is given in section V.

**Related works**

Black and William (2009) suggested that providing feedback is essential in formative assessments to advance learners forward. Students receiving feedback may take steps to remedy whatever weaknesses the assessment exposes (Buchanan, 2000). Maughan (2001) stressed that the feedback of on-line formative assessments provides information that can instantly respond to the learning and teaching process to highlight needed improvements. Scant works have adopted learning feedback on formative assessments (Buchanan, 2000; Henly, 2003; Wang, 2008), and have simply focused on providing predetermined answers as learning feedback. For example, Henly (2003) used a commercial web-based formative assessment system, WebCT (http://www.webct.com), for dental education, in which students can access correct answers immediately after finishing their assessments. Buchanan (2000) developed a web-based formative assessment system, PsyCAL, for a psychology course. An important feature is that for each multiple-choice question answered incorrectly, the system provides a reference (appropriate explanation sections of textbooks) for students to further understand the reasons, instead of directly providing the correct answer. Scholars view feedback interaction as an essential component of formative assessment (Wang, 2007). Pachler et al. (2010) explicitly indicated focal points in the development of on-line formative assessment tools as 1) functions of feedback and 2) sharing outputs and ideas with peers. Costal (2010) stated that providing a direct resource increases the possibility of students assessing online formative assessment. Thus, providing appropriate learning feedback on formative assessments has become a major concern.

Annotations refer to marks readers make on reading materials. Annotations a reader makes will be helpful to subsequent readers (Hwang et al., 2007). Adopting collaborative annotations on e-materials and e-documents has proven to enhance not only learner knowledge, but also their motivation (Hwang et al., 2007; Su, et al., 2010; Yang et al., 2004; Yeh & Lo, 2009). For example, Hwang et al. (2007) proposed a web-based annotation system, capable of collaboratively sharing multimedia annotations on learning materials. Yang et al. (2004) developed a system to provide annotation creation and real-time discussion for collaborative learning. Su et al. (2010) also designed an annotation management system for sharing collaborative annotations on learning documents. However, studies discussing the learning performance and behavior of students when harnessing collaborative feedback in the process of an on-line formative assessment are scant. Certain researchers have admonished, “formative assessment effectiveness across samples and educational contexts is not well known” (Wilson et al., 2011).

This study proposed an online formative assessment tool for the course “Electrical Commerce,” a mandatory course in major and specialist programs, to enhance student willingness to discuss, refer, and rectify their misconceptions on
quizzes of formative assessments, by using collaborative annotations. Annotations and knowledge from peers serve as scaffolding feedback for students, to not only motivate learner interest, but to also reduce frustration (Shute, 2008).

Although most works emphasize a required learning mechanism in formative assessments for students, an efficient management mechanism is also important for teachers to administer formative assessments. Chen and Chen (2009) presented a mobile formative assessment system to help teachers precisely assess individual learning performance, but they did not provide learning feedback for students. Lazarinis (2010) presented an adaptive web testing system for teachers to manage and use different adaptive rules to create formative assessments for students. Testing procedure adaptation relies on performance, prior knowledge, and the goals and preferences of test participants.

This study not only adopts collaborative multimedia annotations as feedback for online formative assessments, but also develops the proposed system to facilitate both student learning and teacher management of formative assessments.

**Proposed approach and system**

**The proposed approach**

The formative assessment quizzes are based on textbook material and handouts. To build the Question Bank Database for formative assessments, teachers have to first build representative questions and their corresponding answers and a question bank for each chapter. This study uses multiple-choice format for most questions.

In the proposed approach (Figure 1), the operation procedure consists of two phases: Test and review phases. Students in the test phase take an on-line formative assessment, which pertains to one chapter of a curriculum. Following the test phase, student learning records are stored in the User Learning Portfolio Database. The procedure then enters the review phase, consisting of the following steps:

1. **Step 1:** After finishing the test phase, students obtain their test results, including their own answers, the correct answers, and the question restatements, to allow them to monitor their knowledge of the subject. A student is supposed to check every incorrectly answered question (IAQ). Comparing the correct answers with the student answers detects these IAQs. The review process in the assessment terminates after the student addresses every IAQ. Otherwise, the student must continue to examine the next IAQ.
2. **Step 2:** When encountering an IAQ, a student reviews it and its corresponding multimedia annotations extracted from the Collaboration Annotation Bank Database. These annotations uploaded by peers are stored in this database.

**Figure 1. Flowchart for the operation procedure**
• Step 3: If the shared annotations are meaningless, the student is encouraged to figure out the whys through surfing the Internet or textbooks, and then annotate it for sharing between peers. To boost student desires to participate, several measures have taken as incentives. First, a Credit Bulletin Board presents to rank the number of given credit each author has gained. Second, a Contribution Bulletin Board presents to rank the number of each author’s annotations. Then the student goes back to Step 1.

• Step 4: If the shared annotations are beneficial, the student directly returns to Step 1 without making any annotation.

Every formative assessment follows the same procedure above. Even after a period, students can still login to review their previous (or historical) formative assessments and the review process follows the review phase (Figure 1).

Providing students with the support of collaborative annotations may facilitate acquiring comprehensive concepts, sustaining connections to the online learning community, and obtaining learning skills to cope with academic problems (Chen & Chao, 2008). Therefore, using our proposed approach, students can take advantage of this cooperative community through the following use scenarios.

Scenario 1: When reviewing the IAQs sequentially, a student finds if there are existing beneficial annotations. If yes, he/she comprehends the whys and further rectifies his/her misconception. If no, he/she figures out the whys by himself/herself through surfing the Internet or textbooks, and then annotates it by copying (and pasting) the Internet address or inputting the textbook location (what page and paragraph). Later, he/she will receive a notification e-mail from the system, which contains a hyperlink linking to the newly inputted annotation.

Scenario 2: A student who does not login to the system for several days will receive an email reminder, which notifies the user to use this system more often. When connecting to the system, the student can observe whether his ranks in the Contribution Bulletin Board and the Credit Bulletin Board are low, thus alerting him that his classmates studied harder than he did. This scenario may also raise his interest to keep up with mainstream studying in the class.

Scenario 1 exhibits how a student contributes and benefits to and from the learning community. With these supports, students realize useful resources and develop their understanding through exploring resources. Scenario 2 demonstrates how such a community raises student awareness of learning progress, which may enable a sustained connection between community members.

Teachers serve as moderators in such an environment to encourage and supervise annotating activities because a lack of moderator feedback would reduce motivation to participate in activities (Chao & Chen, 2009). To prevent inappropriate annotations, the teacher periodically and regularly patrols if the newly inputted annotations are beneficial and then deletes or modifies them if necessary, using the back-end administration platform. The teacher also informs a student of his/her improper annotations through the system so that he/she can modify them online. Students can also report inappropriate annotations to the teacher for further handling. Besides, the questions with high failure rate, annotation quantity, or review time, can be deemed as the challenging or puzzling questions for most students. After analyzing the portfolio of these questions, a teacher can offer guidance by uploading recommendatory annotations and teaching related concepts in class to clarify student misunderstandings. A teacher understands the learning bottlenecks most students face, and further intervenes to offer necessary help. Through the above circumstances, providing an adequate management platform for formative assessments can substantially enhance teaching quality.

The developed system

The structure of the developed system mainly comprises two major parts (Figure 2): (1) Access Database and (2) Web Application, built on Window Server 2003 Internet Information Server (IIS) and written by Microsoft ASP.NET, JAVASCRIPT, and HTML. The Access Database is comprised of a number of relational tables to store the necessary data.

The web application is divided into two parts: the front-end User Platform and the back-end Administrator Platform. The front-end platform is available for users to perform activities on formative assessments, whereas the back-end platform is only available for administrators (teachers) to manage user activities. The back-end Administration
Platform consists of the main management functions: Formative assessment management (FAM), collaborative annotation management (CAM), and other miscellaneous functions including summative assessment management (SAM), remote user management (RUM), and statistic analysis (SA). The striking manipulation and the corresponding GUIs are presented as follows.

**FAM**

The FAM generates web pages for the test content of the formative assessment and identifies student performance and learning barriers (the questions that most students failed). Figure 3 demonstrates some snapshots in which a teacher configures a formative assessment. The upper-left window (Step 1) sets the scopes of a formative assessment and the upper-right window (Step 2) shows the setting of its name, designated class, start time, and end time (Step 2). The bottom window shows the corresponding test content of the formative assessment, which students are going to take it. All students take the same quiz, but the questions are arranged randomly to prevent cheating.

*Figure 3. Snapshots of the administrator setting a FA and students taking it*
The CAM allows a teacher to administer collaborative annotation activities, such as modifying or deleting improper annotations. After finishing a formative assessment, a student can review the results, input and view annotations, receive a notification e-mail when there is a newly inputted annotation, or review historical records. Figure 4 demonstrates some snapshots. The upper-left window shows the historical list of formative assessments that a student has taken, whereas the upper-right window shows the results of a formative assessment. The bottom window shows the composition of a new annotation for a question using the text editor, embedded by a free software component FCKEditor (http://www.fckeditor.net). Using the Edition Tool Set, users can upload multimedia annotations by inputting text, images, tables, flash objects, and hyperlinks in the Edition Area and further arrange the layout.

![Figure 4. Snapshots of viewing results and editing multimedia annotations](image)

Figure 4. Snapshots of viewing results and editing multimedia annotations

![Figure 5. Example of collaborative annotations corresponding to a question](image)

Figure 5. Example of collaborative annotations corresponding to a question
Figure 5 shows the annotations that contribute to a question, where the author list (Contribution Bulletin Board) is shown in descending order according to the number of annotations. Students who are industrious to generate the most annotations are placed at the top of the author list. Users can select a desired author or all authors to read his/her or all annotations, respectively, which are displayed at the bottom. If users feel that an annotation is beneficial, they can give credit to the author. Users can notify the teacher of any improper annotation for further handling.

**Miscellaneous functions**

Further miscellaneous functions include SAM, RUM, and SA. The SAM is similar to FAM, but for summative assessments. The content refers to that of formative assessments, but has slight modifications. Each student receives the same summative assessment questions, but in random order. The RUM is in charge of managing the user accounts, whereas the SA is in charge of computing the statistics (grades) of the formative and summative assessments.

**Educational evaluation**

**Objectives**

This evaluation compares the proposed system – the Collaborative Annotation Formative Assessment System (henceforth referred as CAFAS) – with the system without collaborative annotations – the Non-Collaborative Annotation Formative Assessment System (henceforth referred as NCAFAS). Upon completion of a formative assessment, both systems give students feedback about their performance, including questions, their answers, the correct answers, and scores. However, with CAFAS, students are able to input and review their collaborative annotations, but NCAFAS students have no such support. To conduct the evaluation, the NCAFAS must be developed. The modular design of the developed CAFAS simplifies building the NCAFAS with only slight modifications to the inner software structure of CAFAS. Building the NCAFAS requires simply removing the related modules, for example, the CAM module in the Administration Platform and the CA activity module in the User Platform in Figure 2.

This study was administered to three 3rd-year classes of the Department of International Business in Ching Yun University, with an approximate enrollment of 180 students. The first class, consisting of 53 students, was assigned as the experiment class and used the CAFAS. The second class, comprising 60 students, was assigned as the control class and used the NCAFAS. The third class, consisting of 51 students, was only assigned to validate and analyze the content reliability of the pretest and posttest.

The current study addresses the following issues: (1) Whether CAFAS, compared to NCAFAS, enhances learning achievements and motivation; and (2) Analyzes the behaviors of annotating and reviewing within the experiment class. Finally, the questionnaire and interview were conducted in the experiment class.

**Research tools**

The “Electronic Commerce” course includes seven chapters: “Basic Network Introduction,” “Basic Electronic Commerce Introduction,” “Basic Electronic Business Introduction,” “The Strategy and Method of the Electronic Business System,” “The Plan and Design of the Electronic Business System,” “The Implementation of the Electronic Business System,” and “The Maintenance of the Electronic Business System.” These chapters were taught in both classes by face-to-face lectures through textbook and PowerPoint material. The test content of any formative assessment originates from teaching materials, ranging from those that focus on knowledge and comprehension to those that focus on application and analysis. To establish substantial question banks, two teachers cooperated to collect and edit the question banks, generating 66, 66, 125, 119, 85, 90, and 50 multiple-choice questions for Chapters 1 to 7. A typical question may address a specific term, for example, “What is the phenomenon of the Long Tail,” followed by four possible options, with only one correct option.

Formative assessment quizzes directly reflect summative assessments of the course. Thus, students using the system become familiar with the type of questions they are required to answer on their midterm and final exams. Formative
and summative assessments have few overlaps. Even though similar types of questions appear on both assessments, the questions in a summative assessment are not directly copied from formative assessments, thus eliminating the problem of students memorizing questions and answers. The midterm test content ranges from Chapters 1 to 4, and that of the final ranges from Chapters 1 to 7. The two classes received the same test content.

Formative assessments mainly serve to motivate students to extensive and deep study and to further engage them in annotating in the experiment class. Therefore, this work does not report the test scores in all formative assessments in the following sections.

Research procedure

This research adopted a quasi-experimental design method. Both classes were taught how to use their designated systems at the onset. The experimental treatment lasted for one semester for two hours each week. Both classes adhered to the same schedule. Upon completion of each chapter, both classes received the same formative assessment of the chapter. After finishing one formative assessment, both classes reviewed their results of the assessment. However, students in the experimental class performed annotating activities to fulfill remedial learning. The two classes were encouraged to not only review the learning materials, but also their historical formative assessments before the summative assessments. The midterm and final, taken during the middle and end of the semester, were administered as summative assessments to individually evaluate the learning achievements of the two classes.

To verify possible significant difference in the prior knowledge of students in both classes, a study regarding prior knowledge was conducted before the evaluation. To assure pretest validity and reliability, 2 experts reviewed the content of the pretest, which was then tested by 51 students in the third class. After that, inappropriate questions were removed according to the corresponding difficulty and discrimination levels, resulting in 38 multiple-choice questions and Cronbach’s $\alpha$ being 0.83.

Table 1 shows the results of the independent samples $t$-test, showing no significant difference in the average scores of the basic learning backgrounds between the experiment and control classes ($t = -0.62, p > .05$). The basic backgrounds of the students in the both classes showed no significant difference.

<table>
<thead>
<tr>
<th>Classes</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>53</td>
<td>36.11</td>
<td>13.13</td>
<td>-0.62</td>
<td>0.53</td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>37.81</td>
<td>12.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The validity and reliability analysis of both summative assessments were handled the same as that of the pretest, resulting in 62 and 82 questions, with Cronbach $\alpha$ at 0.86 and 0.88, respectively.

Data collection and analysis

To analyze participant preference, both systems recorded participant activities as logged data, including login time, activity types (reviewing, annotating, or testing), source IP (locations, such as home or school), and stay period (the time a visitor spends on the system). The SPSS was used to conduct the statistical analysis.

Comparison between the experiment class and control class

The independent samples $t$-test was chosen to determine if there was significant difference in the different measures between the experiment class and the control class. The results are shown in Table 2.

The mean score of the midterm of the experiment class (mean = 61.74) is significantly higher than that of the control class (mean = 52.48) ($t = -3.50, p < .05$). Similarly, the mean score of the final of the experiment class (mean = 55.55) is also significantly higher than that of the control class (mean = 47.93) ($t = -3.06, p < .05$). These two identical
results unveil that the use of CAFAS facilitates peer interaction in a collaborative learning context, allowing students in the experiment class to obtain shared annotations and to brainstorm. Therefore, they obtained deeper understanding of online questions in formative assessments and acquired enhanced learning achievements.

The mean of the review time before the midterm of the experiment class (mean = 104.94) is significantly higher than that of the control class (mean = 10.92) ($t = -6.74, p < .05$). Similarly, the mean of the review time between the midterm and the final of the experiment class (mean = 180.43) is also significantly higher than that of the control class (mean = 46.77) ($t = -3.65, p < .05$). These two identical results also reveal that students in the experiment class obtained more aids and resources from the collaborative annotations. Compared to students in the control class, students in the experiment class spent more time annotating, discussing, and sharing their opinions on a formative assessment.

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>52.48</td>
<td>14.98</td>
<td>-3.50</td>
<td>0.001</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>61.74</td>
<td>12.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Exam Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>47.93</td>
<td>11.55</td>
<td>-3.06</td>
<td>0.003</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>55.55</td>
<td>14.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review time before the Midterm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>10.92</td>
<td>43.00</td>
<td>-6.74</td>
<td>0.000</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>104.94</td>
<td>93.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review time between the Midterm and the Final</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>46.77</td>
<td>184.80</td>
<td>-3.65</td>
<td>0.000</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>180.43</td>
<td>201.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison within the experimental class

The paired-samples $t$-test was chosen to determine if there was significant difference in the different measures within the experiment class. The results are shown in Table 3.

For annotation quantity, the mean between the midterm and the final (mean = 7.42) is significantly higher than that before the midterm (mean = 3.98) ($t = -3.37, p < .05$). Similarly, the mean of review time between the midterm and the final (mean = 180.43) is also significantly higher than that before the midterm (mean = 104.77) ($t = -3.74, p < .05$). This is because the students were unfamiliar with the CAFAS at the beginning and required a period to become acquainted with it. With the passing of time, students learn how to share collaborative annotations and receive more benefits from them. As the number of annotations increases, more substantial annotations are generated to meet student requirements. This study originally thought that learning achievement on the final should be obviously higher than that of the midterm. However, oppositely, the mean score of the midterm (mean = 61.74) is significantly higher than that of the final exam (mean = 55.55) ($t = 3.23, p < .05$). This strange phenomenon may be due to the larger scope of the final. Thus, it is much more difficult for students to prepare for the final exam.

<table>
<thead>
<tr>
<th>The Pair of Annotation Quantity</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the Midterm</td>
<td>3.98</td>
<td>53</td>
<td>8.52</td>
<td>-3.37</td>
</tr>
<tr>
<td>Between the Midterm and the Final</td>
<td>7.42</td>
<td>53</td>
<td>12.92</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Pair of Review time</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the Midterm</td>
<td>104.94</td>
<td>53</td>
<td>93.11</td>
<td>-3.74</td>
</tr>
<tr>
<td>Between the Midterm and the Final</td>
<td>180.43</td>
<td>53</td>
<td>201.84</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Pair of Score</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm Score</td>
<td>61.74</td>
<td>53</td>
<td>12.84</td>
<td>3.23</td>
</tr>
<tr>
<td>Final Exam Score</td>
<td>55.55</td>
<td>53</td>
<td>14.81</td>
<td></td>
</tr>
</tbody>
</table>

A Pearson Correlation was employed to analyze the correlation between the final score, total annotation quantity, and total review time within the experiment class. A significant positive correlation also exhibited between the total annotation quantity and the total review time (the correlation = 0.575, $p < .01$), and between the total review time and the final score (the correlation = 0.429, $p < .01$). Finally, a significant positive correlation demonstrated between the final score and the total annotation quantity (the correlation = 0.337, $p < .05$). Thus, these three variables have a
moderate positive correlation with each other. Therefore, it is reasonable to infer that more annotations or more review time is conducive to better learning achievement.

**Questionnaire and interview**

To understand student satisfaction in terms of their concerns, a questionnaire with a Likert scale ranging from 5 (strongly agree) to 1 (strongly disagree) was given to the experiment class at the end of the study. This questionnaire was based on Su et al. (2010) and further modified to evaluate student feelings about using the CAFAS. Among 53 students in the experiment class, 51 valid questionnaires were collected and used for the data analysis. After completing the questionnaire survey, short interviews were conducted for 9 selected students to elicit their actual feelings.

<table>
<thead>
<tr>
<th>No</th>
<th>Question</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Does the CAFAS provide a convenient environment (friendly user interfaces)?</td>
<td>4.21</td>
<td>0.73</td>
</tr>
<tr>
<td>2.</td>
<td>Does the CAFAS have good stability (quick response)?</td>
<td>3.75</td>
<td>1.11</td>
</tr>
<tr>
<td>3.</td>
<td>What is your overall satisfaction degree toward the CAFAS?</td>
<td>4.00</td>
<td>0.89</td>
</tr>
<tr>
<td>4.</td>
<td>Does the CAFAS aid you in your learning in Electronic Commerce?</td>
<td>3.85</td>
<td>0.91</td>
</tr>
<tr>
<td>5.</td>
<td>Does the CAFAS stimulate you to spend more time on the system?</td>
<td>3.86</td>
<td>1.01</td>
</tr>
<tr>
<td>6.</td>
<td>Does annotating impose an additional load?</td>
<td>3.95</td>
<td>1.04</td>
</tr>
<tr>
<td>7.</td>
<td>Do you look into the questions that have the highest review time?</td>
<td>3.69</td>
<td>1.03</td>
</tr>
<tr>
<td>8.</td>
<td>Do you look into the questions that have the most annotations?</td>
<td>3.79</td>
<td>1.05</td>
</tr>
</tbody>
</table>

The questionnaire results, shown in Table 4, reveal positive feedback for most of the evaluated aspects. The results of questions 1 to 3 show that most students were satisfied with the CAFAS and agreed that it is a stable and convenient system. For example, one interviewee stated, “The user interfaces are friendly. It is easy to use even for the first time.” Conversely, there was also one opposite opinion: “Sometimes the system responded slowly. However, it may relate to the low network bandwidth of my home PC.”

The results of Question 4 show that most students also agreed that the CAFAS is a practical auxiliary tool for learning Electronic Commerce. Most students generally agreed that the support of shared annotations was useful in providing different aspects of knowledge. For example, one interviewee stated, “Some recommended annotations and key sentences provided by other peers actually helped me to better understand the questions. Sometimes I pondered the meaning of the annotated sentences and content.” One interviewee stated, “I did not have to review the questions that I understood before the exams so that I could focus on reviewing incorrectly-answered questions on my historical formative assessments.” Conversely, 3 interviewees simultaneously expressed that they preferred reviewing the materials through papers or textbooks rather than using computers before summative assessments. One of the 3 devised an approach to cope with such a dilemma, saying, “Before the exams, I copied the content of my historical formative assessments and pasted it into Microsoft Word, and then attached useful annotations to the corresponding questions. After proper arrangement, I printed them so that I could read them in my own way.”

The results of Question 5 show that the system moderately stimulated students to spend more time on it. Six of 9 interviewees expressed that they often reviewed or annotated at home. One stated, “I spent more time on the system, especially before the day of summative assessments because it provided sufficient question banks and thus became my important resource to practice.” One stated, “I enjoy praise when someone gives credit to my annotations and this inspires me to contribute more.” Another interviewee commented, “The collaborative annotations interest me because they can present in multimedia form, which is much easier to understand.”

Certain students expressed that they prefer independent study, rather than collaboration, whereas others expressed that they prefer their peers as a source of guidance. Some even asserted that they dislike sharing their own knowledge with others. A few students stated that they were not enthusiastic about participating in collaborative discussion. Accordingly, we speculated that learning styles (Huang et al., 2009), including independent, dependent, competitive, collaborative, avoidant, and participant, may also play a crucial factor in annotating and reviewing behaviors in formative assessments and such speculation needs further verification in the future.
When investigating whether annotating imposes an additional load (Question 6), 16 of 53 students (30%) agreed; 28 were neutral (53%), and the other disagreed (17%). Although the students welcomed collaborative annotations, some suffered from an extra load associated with annotating. This may be because (1) some felt there were too many formative assessments; and (2) some felt frustrated with numerous incorrectly answered questions in the formative assessment. One interviewee expressed, “At the beginning, I was interested in a collaborative environment. However, after several months, I felt there were too many questions needing to be addressed and eventually I grew tired of annotating.” Another interviewee stressed, “I was certainly frustrated by too many incorrectly answered questions and then I reluctantly annotated or even skipped them.”

From the results of questions 7 and 8 in the questionnaire, we infer that questions with more review time or more annotations are the topics of most concern to students or that confuse many students. A teacher can see the corresponding annotations of students and further assess their level of comprehension for instructional intervention. Harnessing collaborative annotations on online formative assessments are an effective way for teachers to identify most students’ hurdles. Thus, the teacher can spend more time explaining these points to further clarify most student misunderstandings. Questions with more annotations may encourage students to look into and discuss more details. One interviewee remarked, “I was curious about a question that had been annotated by many peers, because I wanted to know their arguments. I also added my viewpoints if necessary.” Another interviewee stated, “I looked into a question with many annotations for more cues when I did not understand its meaning or its technical terms.”

Conclusion

This research adopted collaborative annotating on formative assessments to boost learning efficiency. The Administration Platform of the developed system CAFAS enables a teacher to organize formative assessments and to manage annotation activities, whereas the User Platform allows students to conduct and review formative assessments and to annotate collaboratively.

The comparison between the experiment class (with CAFAS) and the control class (with NCAFAS) shows that the mean scores of the midterm and final of the experiment class are both significantly higher than those of the control class. These results reveal that using the CAFAS enhances learning achievement. This finding is consistent with that of Hwang et al. (2007), who stated that the use of shared annotation on e-material enhances learning achievement. In addition, review time before the midterm and between the midterm and final in the experiment class are both significantly higher than those of the control class are. These results show that the use of the CAFAS moderately enhances student motivation to spend more time on the course.

Within the experiment class, the means of annotation quantity and review time between the midterm and the final are both significantly higher than those before the midterm. These results show that students are willing to devote more annotations and consider the CAFAS as an important learning resource when they gradually acquire more substantial aids through collaborative annotations. The results of correlation analysis exhibit three variables: annotation quantity, review time, and the final score have positive correlations.

The results of the questionnaire and interview reveal that most students have positive attitudes toward the CAFAS. They agree that the CAFAS is a convenient and stable environment. Such a system moderately stimulates students to spend more time on it and aids their learning in Electronic Commerce. The questions with much review time or many annotations can serve as diagnostic cues for a teacher in exploring students’ learning context and in turn adjusting instruction.

To establish such a system, teachers must first pay some efforts on generating multiple-choice questions at the beginning. However, an established quiz database can persistently afford current and prospective students the opportunity to benefit from the system. The current system only provides the user interface for teachers to manually input questions. Future works should develop the helpful function of generating multiple-choice questions and choosing distracters by applying Article Intelligence (AI) techniques. Thus, the burden of teachers in generating questions and distracters can significantly reduce.
References


